Non linear frequency compression with particular reference to helium speech.

Al-Sulaifanie, Bayez K.

Award date:
1984

Awarding institution:
University of Bath

NON LINEAR FREQUENCY COMPRESSION WITH PARTICULAR REFERENCE TO HELIUM SPEECH

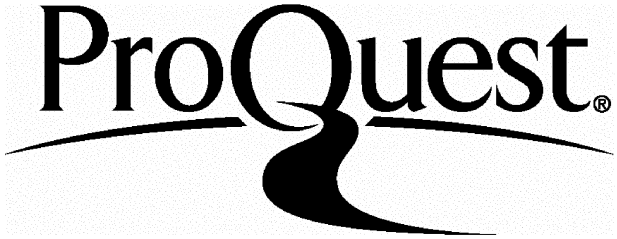
Submitted by

BAYEZ K. AL-SULAIFANIE

for the Degree of PhD

of the University of Bath

1984



To little Joanah

CONTENTS

	SUMMARY	v
	ACKNOWLEDGEMENTS	vi
	CHAPTER ONE - INTRODUCTION	1
1.1	Speech Translation	1
1.2	Efficiency of Speech Transmission	1
1.3	Speech Compression Systems	5
1.4	Non Linear Frequency Compression	7
	1.4.1 Deep sea diving	8
1.5	Conclusion	12
	CHAPTER TWO - SPEECH PRODUCTION MODELLING	13
2.1	Introduction	13
2.2	Speech Production	13
	2.2.1 General	13
	2.2.2 The speech waveform	15
	2.2.3 The vocal tract transfer function	19
	2.2.4 The sources transfer function	25
	2.2.5 The radiation transfer function	26
2.3	Speech Perception	27
2.4	Helium Speech Production	29
	2.4.1 Formant frequencies in Helium Oxygen environments	29
	2.4.2 Formant bandwidth	36
	2.4.3 The source transfer function	36
2.5	Conclusion	39

CHAPTER THREE - HELIUM SPEECH UNSCRAMBLERS 40

3.1 Introduction 40

3.2 Time Domain Techniques 40

3.2.1 Recording and play back technique 40

3.2.2 Waveform segmentation and expansion 41

3.2.3 Autocorrelation subtraction technique 41

3.3 Frequency subtraction technique 43

3.4 Vocoder technique 43

3.4.1 Hustle 44

3.4.2 Formant restoring vocoder (FRV) 47

3.4.3 Voice Transcoder 47

3.5 Analytical Signal Rooting Technique 49

3.6 Fast Fourier Transform Technique 52

3.7 Convolution Technique 53

3.8 Linear Predictive Technique 55

3.9 Conclusion 57

CHAPTER FOUR 58

4.1 Introduction 58

4.2 Principle of Analysis-Synthesis 59

4.2.1 Short time fourier transform 60

4.3 An Analysis-Synthesis Technique for Helium Speech Processing 66

4.3.1 Relationship between vocal tract transfer function of Helium and Normal Speech 67

4.3.2 Relationship between the radiation characteristics of Helium speech and normal speech 69

4.3.3 Relationship between the source transfer function of Helium and Normal Speech 70

4.3.4 Relationship between the transfer function of the speech production model of Helium Speech and Normal Speech 71

4.3.5 Relationship between the short time fourier transform of Helium Speech and Normal Speech 74

4.3.5.1 The relationship between the instantaneous frequency of Helium Speech and Normal Speech 78

4.3.5.2 The relationship between the envelope of Helium Speech and Normal Speech 81

4.4 Conclusion 83

CHAPTER FIVE - THE REAL TIME ENHANCEMENT SYSTEM 84

5.1 Design Philosophy 84

5.2 The Analysis Section 86

5.2.1 Bandpass filters 86

5.2.2 Envelope detectors 88

5.2.3 Zero crossing detectors 88

5.3 The Processor 93

5.4 The Synthesiser 96

5.4.1 Digital sinusoidal synthesiser 96

5.4.2 Amplitude weighting circuit 103

5.4.3 Output accumulator 106

5.5 Timing Signals 108

5.5.1 Analyser Control Signals 109

5.5.2 Up date control signal 116

5.5.3 Synthesiser Control Signals 116

CHAPTER SIX - OVERALL SYSTEM EVALUATION 120

6.1 Introduction 120

6.2 Diagnostic Rhyme Test (DRT) 121

6.3 Helium Speech Evaluation Tests 122

6.4 Spectral Display of the Speech Signal 123

6.5 Steady State Spectral Evaluation 124

6.5.1 Voiced case 125

6.5.1.1 Formant compression 125

6.5.1.2 Amplitude equalisation 130

6.5.2 Unvoiced case 134

6.5.3 Narrow band case 134

6.6 Continuous Speech Testing 137

6.7 Conclusion 140

CHAPTER SEVEN - CONCLUSIONS AND FURTHER WORK 142

7.1 Conclusion 142

7.2 Further Work 144

REFERENCES 146

APPENDICES 151

SUMMARY

Helium speech is a term used to denote the speech produced by a deep sea diver breathing a Helium Oxygen mixture. The replacement of nitrogen in normal air by Helium solves some of the physiological problems associated with diving under pressure. However, it introduces severe distortion in diver's speech. The principal distortion is the nonlinear frequency expansion in the formant frequencies.

A real time enhancement system has been constructed and partially tested. The design specification for this unscrambler has been generalised to enable the system to correct most of the Helium speech distortions. The system operates in the frequency domain and is based on the wide band analysis-synthesis technique.

The system's algorithm for correcting the Helium speech distortion, is flexible and could be easily changed to satisfy different diving conditions. The possible use of the system to study Helium speech characteristics has also been considered.

ACKNOWLEDGEMENTS

The author greatly acknowledges Dr R.J. Holbeche for his constant encouragement and advice during this research. Gratitude is also due for his patience and understanding during the writing of this thesis.

The author is grateful to the staff of the Communication Research Laboratory for their help.

Miss S.C. Hussey for typing this thesis.

CHAPTER ONE

INTRODUCTION

1.1 SPEECH TRANSLATION

Historically frequency translation (compression) of speech signals is a useful method for transmitting the speech signal more efficiently over existing circuits. For analogue transmission this frequency scaled signal must be transmitted within a reduced bandwidth. Whilst for digital transmission this scaled signal would be transmitted at a low bit rate. Another important application of speech translation is to unscramble the distorted speech produced by divers breathing Helium Oxygen mixture in deep sea diving. This type of translation is required to be non-linear to match the frequency distortion of the diver's speech.

1.2 EFFICIENCY OF SPEECH TRANSMISSION

In speech commmunication, it has been recognised for many years that significant mismatch exists between information transmitted through a channel and the ability of a listener to ingest this information. The transmission of unnecessary information increase the information rate, the bandwidth of the transmission, the cost and complexity of the transmission system.

There are several techniques for speech communication: one technique is to preserve the acoustic waveform of the speech signal; an alternative technique is to represent speech according to a Parametric model where only the essential parameters of speech production and perception are preserved. Although the first technique is effective it is inefficient in speech transmission.

The second technique is efficient and its communication quality is acceptable in many practical applications.

One way of demonstrating the possibility of compressing the frequencies of the speech signal is to estimate the information carried by the speech signal and the information capacity of the transmission channel.

In speaking, the information in a message is converted to an acoustic signal via the speech production mechanism. The information transferred by speech can be represented by continuous elements from a finite set of symbols. The symbols from which every sound can be classified are called phonemes. Each language has its own distinctive set of phonemes. English language can be represented by a set of approximately 42 phonemes[1].

The amount of information conveyed by an event (phoneme), according to information theory, is closely related to its probability of occurrence. If x_i set of messages are independent and have probability of occurrence $P(x_i)$ then the information (I) associated with the selection of any member of this set is given by [2]:

$$I = - \log_2 P(x_i) \text{ bits} \qquad 1.1$$

The average information carried by a group in the set is given by:

$$H(x) = -\sum_i P(x_i) \log_2 P(x_i) \quad \text{bits} \qquad 1.2$$

If for simplicity it is assumed that all English phonemes were equiprobable, then the average information per phoneme would be approximately 5.4 bits.

In normal conversation approximately ten phonemes are produced per second[3]. The average information carried in speech at this

rate is 54 bits/sec. This amount of information is approximately equivalent to the information rate carried by a binary code representing the 42 English phonemes (in six binary numbers). So it can be said that in this case the written equivalent of speech contain information at 54 bits/sec.

In a communication channel the maximum error-free rate of information (capacity) carried by a transmission channel is defined by Shannon-Hartley's Theorem[2]. This capacity c, for bandwidth BHz and signal to noise ratio S/N.db is given by:

$$C = B \cdot \log_2 (1 + \frac{S}{N}) \quad \text{bits/sec} \qquad 1.3$$

with a given S/N ratio it is clear that the channel capacity is directly proportional to the speech bandwidth.

To transmit speech with communication quality, the bandwidth requirement is approximately 3kHz with a S/N ratio of approximately 30db. This requires a channel transmission capacity of at least 30 kbits/sec to transmit with smallest error the speech signal. If high fidelity transmission is considered with bandwidth of 6-7kHz and S/N of 60-70db, this capacity would be increased to approximately 300 Kbits/sec.

If the required channel capacity for normal conversation is compared with the capacity of the speech channel it is well in disagreement with the information theorem given by Shannon which states that for a signal with certain information rate there should be a coding method which enables a channel with capacity equal to that rate to be transmitted with tolerable error. In this example the capacity of the channel is 60 to 600 times greater than the

information rate of written speech. Further the estimated bandwidth to transmit a normal conversation (54 bits/sec) with a S/N ratio of 30db is 5.4 Hz which implies a possible bandwidth reduction by approximate factor of 600.

However, the written equivalent approach to normal speaking does not take into account important factors such as identity, emotional state, loudness etc. of the speech, which should increase the information carried by speech signal. However, it is difficult to define the additional information required to convey these factors because in continuous speech this information cannot be measured unless it be represented by a specific code. In speech communication the precise definition of those speech parameters that are both necessary and sufficient to convey the speech information from the speaker to the listener provide fidelity criterion for representing the speech in terms of a specific code.

However, it was established that human beings cannot process speech information in excess of 50 bits/sec [3]. But it has also been established that the listener processes this information in different ways. For example a listener could pay attention specifically to background noise or the voice quality and not concentrate on the actual message content[4]. This could result in a reduction in intelligibility of the perceived speech. Even with this sort of fidelity criterion imposed on speech coding previous frequency compression systems proved that there is still substantial redundant information in the speech signal. Indeed the vocoder invented by Holmer Dudly in 1928[5] proved that compression of the bandwidth of the speech signal is possible. However the resulting speech quality was very poor, it was proved that a limited number of

parameters derived from the speech signal could be accommodated by a fraction of bandwidth needed for normal speech transmission and still be capable of producing intelligible speech.

1.3 SPEECH COMPRESSION SYSTEMS

Most efficient speech compression systems depends on an analysis-synthesis speech technique. Analysis-synthesis is an accurate representation of the human speech production mechanism. The principle is to describe the speech signal in terms of the motion of speech production organs. The slow variation of the vocal tract with times imposes a band limitation on the important parameters of the speech signal, i.e. format frequencies, pitch etc. To reduce the bandwidth, only these parameters are transmitted. At the receiving end they control a synthesiser which simulates the speech production model.

Further bandwidth reduction could be imposed by taking into consideration speech perception characteristics. The ear resolves the speech signal into bands, similar electronically to overlapping band pass filters. This allows splitting of the speech signal into small frequency bands and enables them to be transmitted separately, thus reducing the necessary transmitted parameters.

One of the earliest analysis-synthesis techniques which demonstrated the possibility of reducing transmission bandwidth of speech was the device invented by Homer Dudley in 1928[5]. This frequency compressing device was called a vocoder (voice coder) and represented speech by eleven slowly varying signals. Each one of these signals occupied an approximate bandwidth of 25Hz. The total

bandwidth occupied by the transmitted signal was approximately 300Hz.

The vocoder has led to a generation of devices classified under the same name, certain of which are attempting to achieve further bandwidth reduction by transmitting only the significant channels [6]. Others were aiming to improve the speech quality of the original vocoder[7]. In the vocoder the pitch value and voiced/unvoiced decision is required. The pitch measurement is difficult and inaccurate which usually result in poor speech quality [8].

Another group of vocoder is called the frequency dividing vocoders which avoid the pitch detection. They compress the frequencies of the speech signal by dividing its instantaneous frequency. Various systems that used this technique have been proposed[9, 10, 11, 12], they differ by the method by which they compress the speech frequency. Either the instantaneous frequency of the whole speech waveform is compressed[9] or the speech waveform is split into many frequency bands and the instantaneous frequencies of the signals in these bands are compressed separately[10, 11, 12].

The main reasons for different techniques is the incorrect compression of the spectral component of the speech signal when more than one frequency components are present in the speech signal. When the instantaneous frequency of a complex waveform is divided by a certain factor the dominant frequency component of this waveform will translate correctly by that factor. Whilst the non dominant components will translate downward such that their spacing from the dominant components will remain unchanged in both normal and compressed speech[13,14]. Also the amplitude of the non dominant

frequency components is attenuated relative to the amplitude of the dominant component[13, 14].

However, splitting the speech signal into many bands and processing these bands separately can result in correct compression of different frequency components of the speech signal. Bandwidth limitation can be imposed by subjecting the envelope of the signals in different frequency bands to a rooting characteristic law[11, 12].

One of the most important frequency compressing systems is the phase vocoder[15]. In this type of frequency compressing system the speech signal is divided into narrow frequency bands such that only one harmonic is ever present in each band. These bands are then represented by their short time amplitude and phase spectra. The rate of variation of the phases is divided by a constant factor and the amplitude is left unchanged. Originally this system had been simulated on a computer and a two to one compression ratio achieved with good speech quality. The technique used in the phase vocoder to describe short time amplitude and phase of speech spectra can be used to describe most analysis synthesis technique of the speech-processing.

1.4 NON LINEAR FREQUENCY COMPRESSION

Existing frequency compression techniques could be developed to improve the distorted speech produced by deep sea divers.

1.4.1 Deep Sea Diving

Deep sea diving techniques are required to satisfy special commercial or military operations. In the commercial sectors the demand for deep sea diving grows as the importance of the off-shore petroleum industry increases, whilst military capability for submarine salvages increased in 1924 and 1963 from a depth of 304 FSW to 8400 FSW respectively[16]. The requirement continues to increase.

Initially it was always the ability of the diver to hold his breath which limited his underwater work. Even with the introduction of the diving bell the air supplied was limited by the bell's volume. However, the introduction of air compressors increased the amount of air available to divers but physiological effects have emerged. Fast decompression causes what is known as decompression sickness (the bends) in which bubbles of gasses appeared in the blood stream causing fatal effects. The build up of carbon dioxide produces toxic effects while nitrogen in normal air dissolves under pressure in the blood stream causing narcotic effect or narcosis. The second effect has been overcome by using an absorbing chemical to remove carbon dioxide while the third effect is avoided by replacing the nitrogen in normal air with helium. The decompression problem was solved by limiting the permissible time for which the diver could stay under water. Beyond certain depth the permissible time becomes impractical and saturation diving techniques have been introduced to give divers an unlimited amount of time underwater.

For saturation diving the deep diving system (DSS) consists of a decompression chamber (DCC) and a personal transfer capsule (PTC)[16]. The DSS with the diver inside is pressurised to the equivalent work pressure. Then the diver is transferred to the PTC which is detached form the DSS and deployed to the work site. At the required depth the external and internal pressures are equal and the diver emerges to work. After the task is completed the PTC is mated to the DCC which is then decompressed slowly back to normal air pressure. The careful control of the decompression process and the replacement of nitrogen with helium has largely solved the physiological problems associated with deep sea diving as increased the length of time the diver can stay submerged. However, the helium oxygen mixture causes a severe distortion in diver speech. This is usually known as Helium speech.

Helium speech is extremely difficult to understand. The lack of intelligibility increases with depth and with the proportion of helium in the helium oxygen mixture. The main cause for this distortion is the non linear shift in formant frequencies of the speech. The replacement of normal air with the mixture of helium and oxygen changes the transmission characteristics of sound in the human vocal tract. This leads to linear multiplication of vocal tract resonance frequencies (formants) in helium environments compared to that of normal air. The fact that divers work under pressure different from that of normal air alters the physical characteristic of the vocal tract which causes additional shifts in formant frequencies. The low frequency formants shifts more than the higher ones and lead to an overall non linear shift in formant frequencies.

Another factor which affects the speech intelligibility is the attenuation in the level of the high frequency part of the speech signal. This reduces the amplitude of the high frequency formants relative to the low frequency formants. The fact that the transition from vowel to consonants lies in the high frequency part of the speech spectrum produces even more distortion of the speech[17]. Also the recent discovery of the increase in formant bandwidth [18, 19] could be the reason for nasalization associated with Helium speech.

The loss of intelligibility in divers speech results in poor communication between the diver and the surface and between divers. Good communication is vitally important in dangerous conditions such as that associated with deep diving operation.

In 1962 R.G. Beil [20] introduced the helium speech problem in a letter to the Journal of Acoustical Society of America. Restoring the intelligibility of Helium speech to normality attracted considerable attention. A number of reports have been written to analyse changes in Helium speech characteristics[17, 19, 20, 21, 22, 23, 24, 25, 26] and many enhancement systems had been built or merely proposed[27-35].

In spite of the volume of work that has been done on this area, it is interesting that no really effective technique has yet emerged. It also appears that very little work has been done on the fundamental aspects of Helium speech production. The effect of non linear shift is not yet known[35]. However, some new characteristics had been discovered recently[19, 25, 35] i.e. the

change in formant bandwidth. Also the region of high frequency attenuation is undefined.

The speculation about, and misunderstanding of Helium speech characteristics[31, 36] could be the main reason for the delay in providing a completely successful enhancement system. Normal speech itself is difficult to analyse, also the hazardous conditions under which Helium speech is produced together with the difficulty of obtaining Helium speech test samples are significant factors which complicate the apparent lack of understanding of this phenomenon.

Previous Helium speech enhancement systems lack flexibility in that they are generally optimised for a specific helium oxygen mixture and a certain depth. Virtually all reported systems except one [35] have been built to descramble the speech under certain conditions and had limited or no facility for optimisation under different conditions. The hardware in these sytems are difficult to change, to satisfy different diving conditions or to respond to new discoveries in the speech characteristics[26]. Flexible systems could be used to determine the contribution of different Helium speech parameters on their intelligibility[36]. Without flexibility in the design of Helium speech enhancement systems, the requirement for the descrambler to operate at greater and greater depths are unlikely to be properly met.

During the past twenty years more than fifteen different systems have been reported. They range from simple systems which simply shift the Helium speech spectrum by a constant factor[28], to complex systems which use fast fourier transform to analyse and subsequently resynthesise the speech[35]. With the exception of the

later sytems, all these enhancement systems have been designed to correct formant frequencies linearly and they include no equalization for the attenuation in the high frequency portions of the speech spectrum.

It appears from literature published recently on Helium speech that greater emphasis is now being placed on Helium speech generation[18, 19, 26, 35, 37]. The recent progress in digital techniques for processing speech could possibly lead to a new generation of enhancement systems providing improved intelligibililty of the enhanced speech and therefore increasing diving efficiency[38]. Complex Helium speech enhancement systems are already emerging[35]. Also advancements in speech processing techniques could possibly lead to improvements over the original Helium speech enhancement systems to meet improved overall performance[37, 39].

1.5 CONCLUSION

Although there has been considerable research reported in the field of Helium speech descrambling, no single flexible system has emerged that produces acceptable performance in real time[26].

The majority of proposed system often produce a linear frequency compression characteristics whereas it has been established[22] that speech produced in a helium oxygen atmosphere is subject to a significantly non linear frequency transformation.

Clearly the development of non linear frequency transformation techniques would significantly contribute to the realisation of improved descramblers.

CHAPTER TWO

SPEECH PRODUCTION MODELLING

2.1 INTRODUCTION

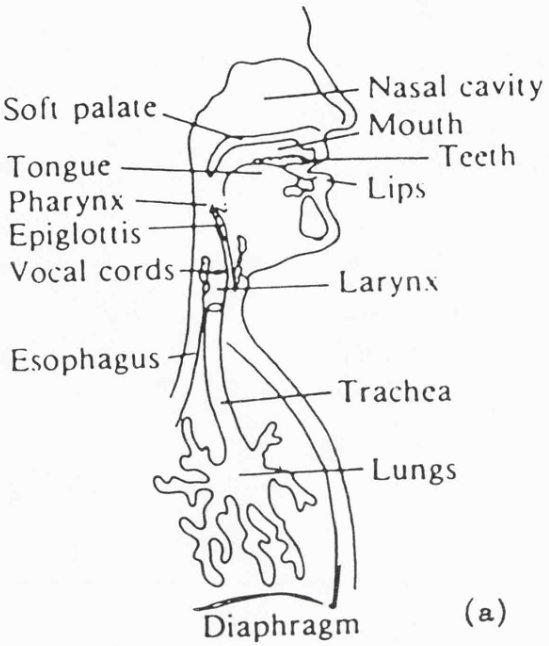
As described in Chapter One frequency compression systems of speech signals rely on the speech production model and the characteristics of the human auditory system. An appreciation of the characteristics of this model is required for the development of improved and more flexible systems. Also it enables the designer to ease the restrictions on many of the design parameters.

2.2 SPEECH PRODUCTION

2.2.1 General

Speech is produced by a stream of air being forced in a controlled way through the larynx and vocal tract. This stream of air excites the vocal tract by the appropriate source, either voiced or unvoiced to produce different sounds. The human vocal mechanism and the representation of its main features are shown in Figure 2.1. The vocal tract represents three main cavities, the pharynx, the mouth and the nasal cavities. In voiced sound the excitation is a train of quasiperiodic pulses representing the air flow through the vocal cords as they vibrate. Whereas for unvoiced sounds the source is generated by air being forced through narrow passages in different parts of the vocal tract thereby creating turbulance, which produces a source of noise to excite the vocal tract.

In electrical engineering terminology the speech production mechanism could be represented by an acoustic filter (vocal tract) excited by two types of sources. The periodic source would be used



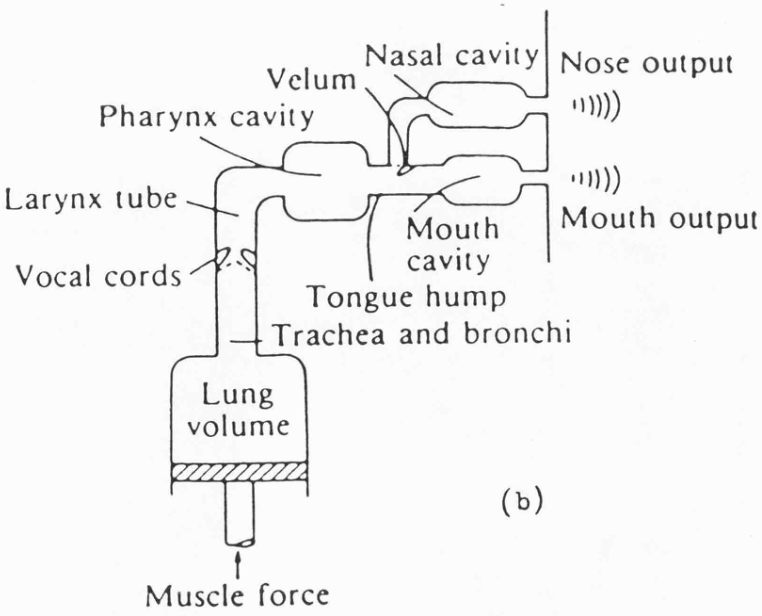


Figure 2.1 Human Speech Production Mechanism

(a) Human Vocal Mechanism

(b) Acoustical Representation of the Vocal Mechanism

to represent the vocal cord vibration and the noise source to represent a sudden release of, or interruption in the air in the vocal tract. The acoustic filter would modify the characteristic of the sources by accentuating certain frequencies and attenuating others. As for any acoustic cavity, the length, shape and medium determine the characteristics of the vocal tract. The vocal tract is a multiresonant frequency filter with its resonance at the formant frequencies. These formant frequencies change with time as the vocal tract changes shape in order to produce different sounds. The variation of these formants with time determine the speech intelligibility.

The rate at which the vocal cord vibrates is known as the speech pitch, this pitch varies relatively slowly and its variation determines the intonation pattern in speech which is important for the naturalness of the sound[40].

2.2.2 The speech waveform

From the above discussion the speech waveform in the time domain could be considered to result from a linear time varying system, the vocal tract, with appropriate excitation. If the vocal tract shape is fixed the output of the system is the convolution of the excitation and vocal tract impulse response. However, the vocal tract slowly varies with time in order to produce different sounds whilst the output over the short term is still approximated by the convolution of the excitation and vocal tract impulse response. The interval during which the speech signal could be considered stationary is between 10-30 milliseconds[3]. The above model shown in Figure 2.2a illustrates the time domain response and spectra of a

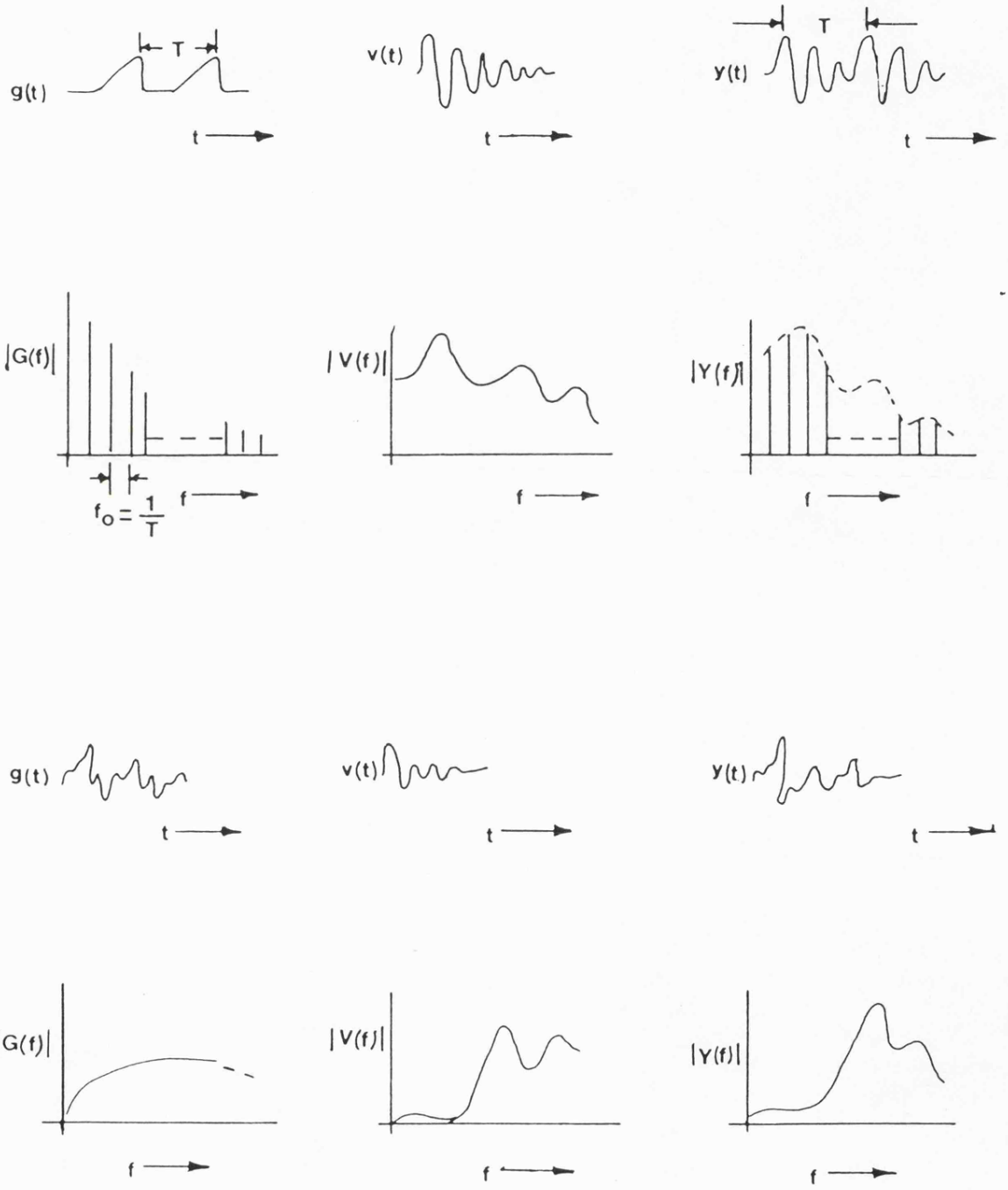


Figure 2.2 Model of Speech Production Mechanism as the response of a quasi-stationary linear system.

(a) Time and Frequency Domain Characterisation for voiced sound.

(b) Time and Frequency Domain Characterisation for unvoiced sound.

g : Excitation , v : Vocal tract , y : Output sound.

segment of voiced speech. Considering the system in the frequency domain, the fourier transform of the speech waveform is the product of the fourier transform of the excitation and the vocal tract frequency response as shown in Figure 2.2a.

Specifically, for periodic excitation the speech waveform is a line spectrum with harmonics spaced at 1/T HZ. The envelope of this spectrum reflects the glottal pulse shape and frequency response of the vocal tract, which is a smooth function of frequency with peaks corresponding to its formant frequencies. In other words the spectrum consists of the product of the line spectrum due to the excitation, and an envelope which characterises the vocal tract transfer function. When the vocal tract changes shape with time to produce different sounds the envelope also changes, similarly, as the excitation period changes for voiced sounds the spacing of the pitch harmonics will change also.

In addition to the effect of the glottal pulses and the vocal tract on the output spectrum the transmission characteristics from the mouth also modify this spectrum. In a similar manner, the unvoiced sounds are excited from noise source which is relatively flat spectrally (Figure 2.2b).

A more realistic model for speech production is shown in Figure 2.3. Here the vocal tract is represented by a four terminal network, with its output connected to the radiation impedance. The network driving the vocal tract represents the source and its impedance.

The transfer function of this system, in terms of its physical parameters is the ratio of the acoustic pressure (voltage) produced

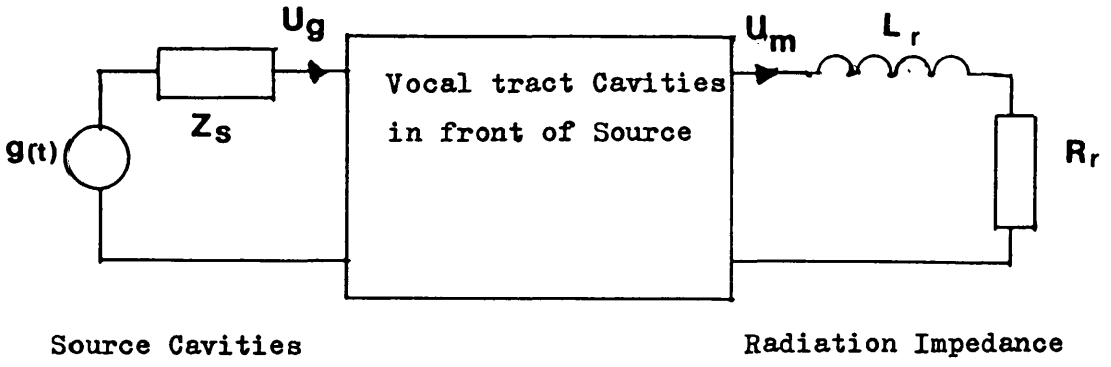


Figure 2.3 Representation of Speech Production Mechanism by Four Terminal Network.

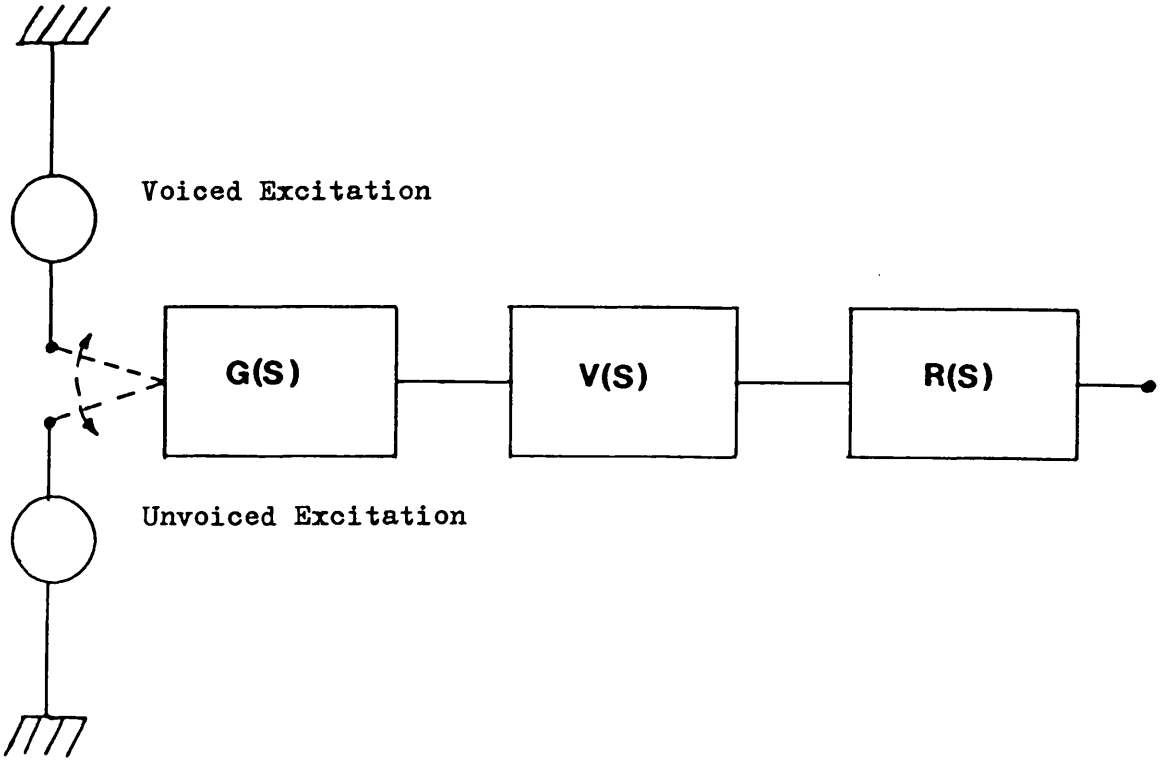


Figure 2.4 Schematic Representation of the Production of Sounds.

at a point in front of the mouth and the volume velocity of the sound source (current). The model shown in Figure 2.3 could be replaced by that shown in Figure 2.4. The transfer functions G(S), V(S) and R(S) represent the sources, the vocal tract and the radiation functions respectively and will be defined later. Mathematically the transfer function of the speech production organ is given by:

$$T(S) = G(S).V(S).R(S) \qquad 2.1$$

2.2.3 The vocal tract Transfer Function

The vocal tract transfer function is the most important part of speech production model because the intelligibility of speech depends greatly on the changes in this function with time. Accurate representation of the vocal tract is difficult because it is a non-uniform acoustic tube. It changes shape with time and is terminated by a variable load (the mouth). Also it is excited by different types of sources with variable internal impedance (glottal impedance).

A simplifying assumption is to consider the vocal tract as a uniform tube which could be represented by a single transmission line. This model is useful in describing characteristics of the vocal tract analytically and it has been used as a basis for designing many speech synthesisers[41, 42]. It will be considered briefly in this section. More vocal tract representation techniques are described in detail by Flanagan[3], 1972 or Fant[40], 1960.

The vocal tract from glottis to mouth may be considered as a uniform lossless tube (Figure 2.5). It may also be assumed that the

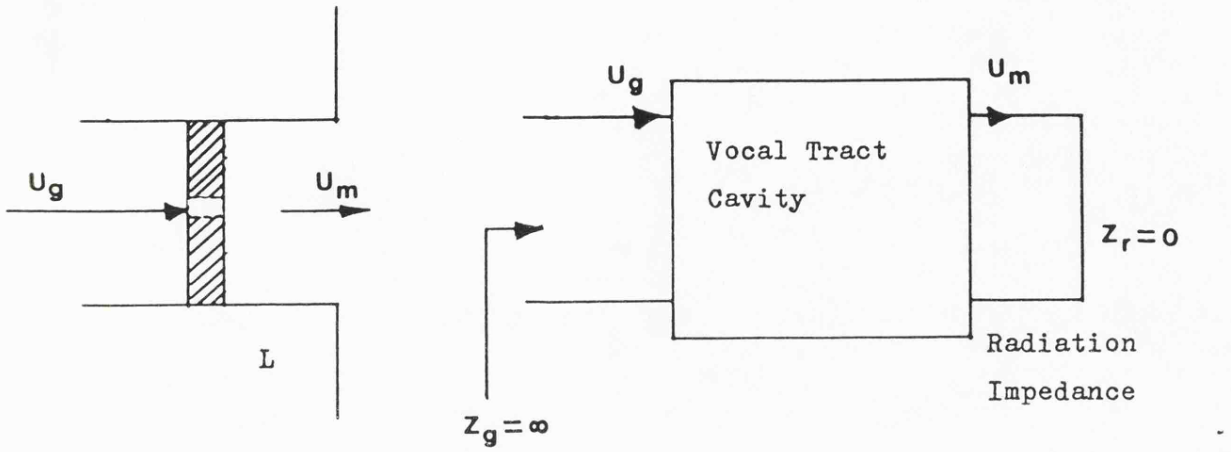


Figure 2.5 Single Transmission Line Approximation to a vocal tract.

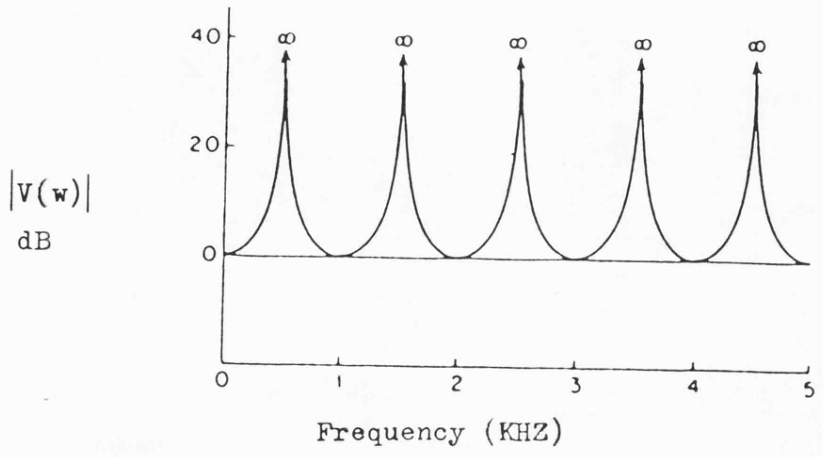


Figure 2.6 Frequence Response for the uniform Lossles Transmission Line Shown in Figure 2.5
(After Rabiner and Schafer (1))

radiation load is small compared with the characteristic impedance of the tract and further that the coupling impedance at the glottis point is very large. An introduction of these assumptions into the equivalent circuit of Figure 2.3 leads to a representation of the volume velocity (Ug) by a current source. The radiation impedance may be replaced by a short circuit which carries a current equivalent to the volume velocity (Um). Then the transfer function of this model can be given by:

$$\frac{U_m}{U_g} = \frac{1}{\cos(\frac{wl}{c})} \qquad 2.2$$

where c is the sound velocity and l is the vocal tract length.

Equation 2.2 has an infinite number of peaks as shown in Figure 2.6. These peaks (poles) represent the formant frequencies F_n, which are given by:

$$F_n = \pm \frac{2n + 1}{2l} \pi c \qquad n = 0,1,2 \qquad 2.3$$

The lossless model described by equation 2.2 is only approximate because of different internal losses associated with the vocal tract and because of the glottal radiation effects.

To modify equation 2.2 to take account of the losses, it can be factorised into an infinite number of conjugate poles in the complex frequency plane as:

$$V(S) = \sum_{1}^{\infty} \frac{A_n S_n S_n^*}{(S - Sn)\ (S - Sn^*)} \qquad 2.4$$

where $S_n = \sigma_n + jw_n'$ are the complex frequency poles and A_n is the formant amplitude. The formant frequency F_n' and

bandwidth B_n are related to σ_n and w_n' by the following equations:

$$W_n' = 2\pi F_n'$$

and 2.5

$$\sigma_n = -\pi B_n$$

In equation 2.5, the effect of internal losses and radiation glottal effects are included in F_n', B_n and A_n. For the lossless case $\sigma_n = 0$ and s_n in equation 2.4 must be replaced by jw_n.

Three types of internal losses are associated with the vocal tract. The friction losses are a result of friction between air and the walls of the vocal tract. The thermal losses are the results of heat conduction from the vocal tract walls. Finally, the wall vibration loss is a result of the change in air pressure as the vocal tract moves to produce different sounds. This vibrates the walls of the vocal tract and increases the energy losses.

The net effect of these losses together with glottal and radiation effects are to shift the formant frequencies downward while broadening the formant bandwidth. The contribution of various losses to formant bandwidth is shown in Figure 2.7, while its effect on the formant frequencies is shown in Figure 2.8. Both internal and glottal radiation effects are functions of the environment under which speech is produced. These effects have been explained in detail in a number of references [1, 3, 35].

Practically the first three of these formants are the most important elements of speech intelligibility. The range of variation of these frequencies for a male subject are[40]:

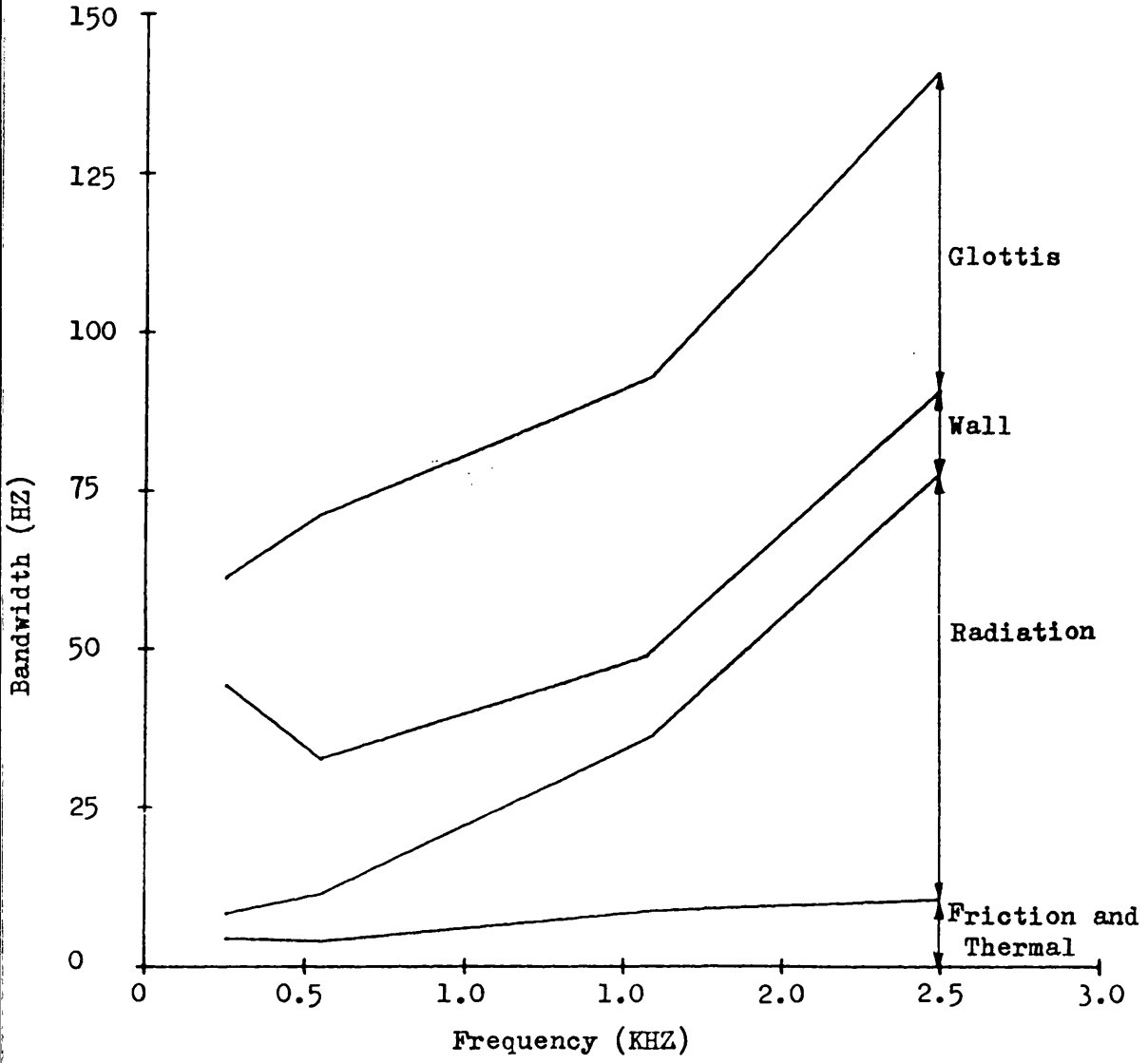


Figure 2.7 Contribution of Various Losses to the Formant Bandwidth in a Normal Atmosphere.
(After Richard (35))

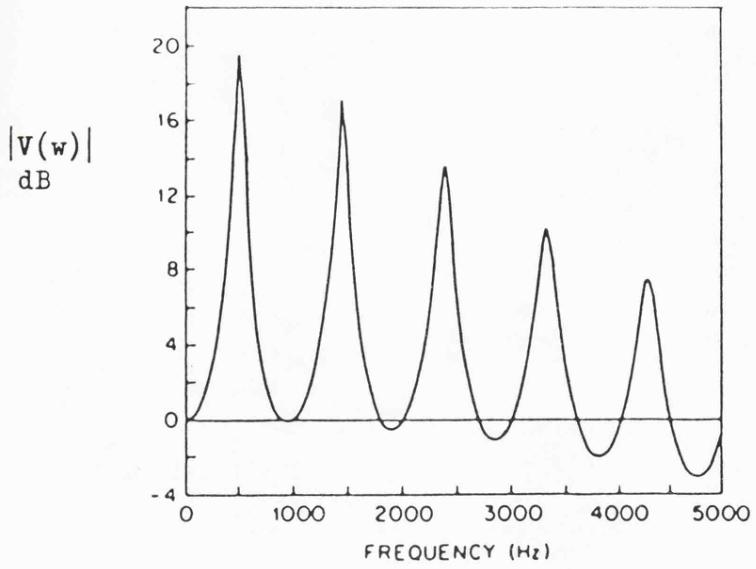


Figure 2.8 Frequency Response for the Lossy uniform Transmission Line Shown in Figure 2.5.
(After Rabiner and Schafer (1))

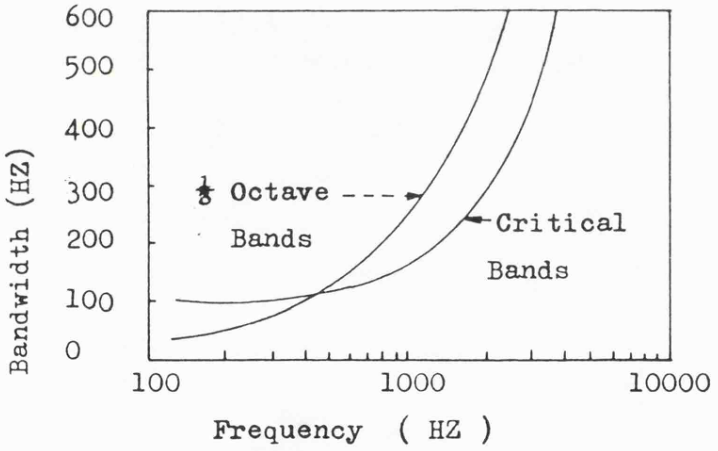


Figure 2.9 Critical Bandwidth as a Function of Critical Centre Frequency.
(After Flangan and Christensen (45))

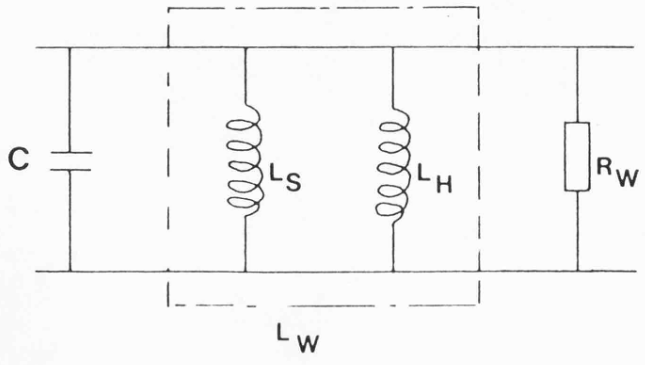


Figure 2.10 Equivalent Circuit of Vocal Tract with total Inductance L_W of the Vocal walls.

First formant: 150-850 Hz.

Second formant: 500-2500Hz.

Third formant: 1500-3500 Hz.

Because females and children have small vocal tracts their formant frequencies are higher. On average females produce 17% higher formant frequencies and children of 10 years of age produce on average 25% higher formant frequencies than adult males.

The bandwidth of the formants is not very important in speech perception providing it is not very large . However, large formant bandwidths introduce nasality effects in speech sounds. The bandwidths of the first three formants are[43]:

B_1: 45 - 130 Hz

B_2: 50 - 190 Hz

B_3: 70 - 260 Hz

2.2.4 The sources transfer function

There are two types of sources which excite the vocal tract, they are the voiced source and unvoiced source. The voiced source is triangular in shape with a duty cycle between 0.3 and 0.7 and spectra relatively rich in harmonics with its power spectrum decaying at 12 db/oct[3]. This model of the voiced source could be represented by periodic impulses at the pitch frequency and a shaping circuit which decays at a rate of 12db/oct. In the complex frequency plane the shaping circuit represent the source transfer

function and can be given by:

$$G(s) = \frac{D}{s^2} \qquad 2.6$$

The factor D represents the level of glottal pulses in the time domain.

The unvoiced source on the other hand is represented by random noise with a flat amplitude spectrum. In the complex frequency plane the shaping circuit associated with the unvoiced source is given by:

$$G(s) = 1 \qquad 2.7$$

2.2.5 The Radiation Transfer Function

The transfer function R(s) is the ratio of the acoustic pressure at a distance in front of the mouth and the acoustic volume velocity from the mouth. The radiation transfer function R(s) has been given by[35]:

$$R(S) = \frac{.31\rho.c.S}{1.44c + 1.07S} \qquad 2.8$$

where ρ and c are the gas density and the sound velocity in the medium filling the vocal tract. The pole of equation 2.8 has little effect on the spectrum of the speech. Hence, the radiation effect from the mouth introduces a factor of approximately 6 db/oct in the speech transfer function. Combining the effects of radiation with the source characteristics, the overall speech spectrum of a voiced sound decays at approximately 6 db/oct.

2.3 SPEECH PERCEPTION

The perceptual aspects of a speech signal is complicated and difficult to understand[44]. However, there are a number of commonly accepted aspects which play important roles in speech processing systems.

It is known that the ear makes a crude analysis of speech by resolving the spectra of the speech signal into a number of frequency bands, similar to that produced by a bank of filters with a third octave characteristics[45] as shown in Figure 2.9. On a short term basis the ear determines the level of the energy in each frequency band and conveys this information to the brain[46]. The phases of the individual signals on the short term basis has no effect on speech perception.

The frequency bands into which the ear resolves the speech signal are called the critical bands. According to the critical band theory[47], the ear treats signals in different bands independently. If more than one harmonic of a complex signal falls within the same critical band then the ear will not correctly detect the presence of the two separate frequencies[48]. Also the threshold of detecting a signal in noise increases as a function of the noise bandwidth providing this band is less than the critical band. If the noise extends beyond this critical band then the threshold does not increase further. However, the perceptual threshold of a sound in the presence of another sound is a function of their levels[49]. Low level sounds in one part of the spectrum will be masked by high intensity sounds in another part of the spectrum. This masking has the greatest effect if the two sounds

fall in the same critical band. But even outside a critical band significant masking occurs, mainly of high frequency sounds by low frequency ones[4]. This property is of great value in making listeners insensitive to low level background noise or distortion when the wanted signal is of high level.

A listener is highly sensitive to a difference in the frequency or intensity of the sounds presented to him for comparison[3]. However, a listener is relatively incapable of identifying sounds presented to him in isolation. When the average listener is presented with pure tones individually he is generally able to identify only five separate tones[50]. Conversely, in comparative tests a normal listener can distinguish up to 350,000 different tones[3]. The ability of a listener to distinguish between alternatively presented tones is called differential discrimination whilst his abilty to identify separately presented tones is called absolute discrimination.

For consideration of different speech processing techniques it is clear that differential discrimination is beyond the ability of a listener to discriminate between different sounds under normal conditions. However, differential discrimination could represent the upper bound on the resolving ability of the perceptual mechanism and hence could be useful as a limiting design factor in speech compression systems.

The differential discrimination is measured in terms of difference limen (DL) or just noticeable difference (Jnd). Two frequencies or amplitudes will be judged "the same" if they differ by less than the Jnd[51]. The differential discrimination

measurement based on synthetic sounds are summarised[52] in Table 2.1 .

Table 2.1: Just noticeable difference of different sound's parameters.

Parameter	Jnd	
	Frequency	Amplitude
F_1	±20Hz	±1dB
F_2	±50Hz	±5dB
F_3	±75Hz	±5dB
F_0	±1Hz	±0.1dB

where F_1, F_2 and F_3 are the first, second and third formant frequencies, while F_0 is the pitch frequency.

Table 2.1 shows that the Jnd of the fundamental frequency is smaller than that of the formant frequencies, whilst the Jnd of the formant frequencies decrease with an increase in frequency.

2.4 HELIUM SPEECH PRODUCTION

2.4.1 Formant Frequencies in Helium Oxygen Environments

As described in Section 2.2 the formant frequencies depended on

the sound velocity and the losses in the human speech production mechanism. It was shown[35] that an increase in pressure has a major effect on the wall losses and a minor effect of the other losses. However, the helium oxygen mixture has an equal effect on all types of losses. The fact that wall losses affect low freqency formants results in nonlinear shifts in the formant frequencies. The net result is to shift the low frequency formants by a greater factor than the high frequency formants.

Fant and Sonesson[22] (1964) gave the basic theory for the nonlinear shift in formant frequencies. They regarded the vocal tract model given in Figure 2.5 to be equivalent to the circuit shown in Figure 2.10. The inductor with value L_s represents the soft parts of the vocal tract and is not pressure dependent. Also they assumed the resonant frequency produced by the soft part of the vocal tract walls is the lowest resonant frequency of the vocal tract (i.e. when mouth is closed). The resonant frequency of this resonator represents the real resonant frequency of the vocal tract under diving environments.

From circuit theory it could be shown that the resonant frequency of the vocal tract is given by:

$$F_a^2 = F_{Ha}^2 + F_{Sa}^2 \qquad 2.9$$

where F_a is the resonant frequency of the vocal tract in normal air.
F_{Ha} is the resonant frequency of the hard part of the vocal tract in normal air.
F_{Sa} is the resonant frequency of the soft part of the vocal tract in normal air.

In Helium Oxygen environment equation 2.9 could be rewritten as:

$$F_h^2 = F_{Hh}^2 + F_{Sh}^2 \qquad 2.10$$

where h denotes the helium oxygen environments.

The hard wall resonant frequency of the vocal tract F_{Hh} depend on the speed of the sound in the Helium oxygen mixture, and so F_{Hh} is then given by:

$$F_{Hh} = \frac{C_h}{C_a} F_{Ha} \qquad 2.11$$

where Ch/Ca is the ratio of the speed of the sound in Helium oxygen environments, to that in normal air.

Substituting equation 2.11 into 2.10 gives:

$$F_h^2 = (\frac{C_h}{C_a})^2 F_{Ha}^2 + F_{sh}^2 \qquad 2.12$$

and substituting F_{Ha}^2 from equation 2.9 into equation 2.12 gives:

$$F_h^2 = (\frac{C_h}{C_a})^2 \left[F_a^2 - F_{Ha}^2\right] + F_{Sh}^2 \qquad 2.13$$

However F_s is the limiting formant frequency and hence is given by[22]:

$$F_s = \frac{c}{2\pi}\sqrt{\frac{\rho}{LsV}} \qquad 2.14$$

where Ls and V is the inductance and the volume of the vocal tract when the mouth is closed. Also, LsV is dependent on the vocal tract shape and is constant for a certain vocal tract and ρ is the gas density. Therefore the ratio of F_{Sh} to F_{Sa} is given by:

$$F_{Sh} / F_{Sa} = \frac{c_h}{c_a}\sqrt{\frac{\rho_h}{\rho_a}} \qquad 2.15$$

substituting equation 2.15 into 2.13 and denoting $\frac{c_h}{c_a}$ by K_1 and $\frac{\rho_h}{\rho_a}$ by K_2 gives:

$$F_h^2 = K_1^2 \left[F_a^2 + (K_2 - 1)F_{sa}^2\right] \qquad 2.16$$

The term $(K_2-1)F_{Sa}^2$ represents the effect of the pressure on the vocal tract frequencies. It is clear that this term has a major effect on the low frequency formants when F_a is small and little effect on the higher formants. This is the main reason for non linear shifts in formant frequencies. The lowest formant frequency of the vocal tract (F_{Sa}) is estimated to be in the range 120-250 Hz[22, 23]. It was reported that the relationship shown in equation 2.16 is a good approximation to the nonlinear changes in formant frequencies[22, 23, 35, 36]. However, recently speculation has

arisen about defining the shift in formant frequency[26] by equation 2.16.

In saturation diving the depth at which the divers work must be known for physiological reasons as equilibrium in pressure must be maintained between the diving chamber and the pressure at the diving depth. The other known factor is the percentage of Helium and Oxygen in the breathing gas. This is fundamentally necessary in order to keep the partial pressure of the oxygen constant at the required depth. For these reasons it is helpful to write equation 2.16 in terms of the pressure and speed ratio.

The speed of sound in the mixture is given by:

$$C = \sqrt{\frac{\gamma P}{\rho}} \qquad 2.17$$

where γ is a function of the mixture ratio only. By substituting equation 2.17 into equation 2.15, K_1 can be written as:

$$K_1 = \sqrt{\frac{\gamma_h P_h \rho_a}{\gamma_a P_a \rho_h}} \qquad 2.18$$

Also substituting equation 2.18 into equation 2.16 and letting P_a= 1 ATA (Atmosphere absolute) for normal air F_h can be written as:

$$F_h^2 = K_1^2 F_a^2 + \left[\frac{\gamma_h}{\gamma_a} P_h - K_1^2 \right] F_{sa}^2 \qquad 2.19$$

However, P_h can be related to the depth d in feet by[35]:

$$P_h = 1 + 0.02949 \cdot d \qquad 2.20$$

Now the necessary data for equation 2.19 can be derived from equation 2.20 and Table 2.2. For example for a 90% helium, 10% oxygen mixture this function has been plotted for various depths between 200 and 1500 feet (Figure 2.11). In this example F_{Sa} is chosen to be 150 Hz.

Table 2.2 Ratio of properties of Helium-oxygen atmosphere and air.

% He	% O_2	$\frac{c_h}{c_a}$	$\frac{\gamma_h}{\gamma_a}$
0	100	0.95	1.00
10	90	1.00	1.01
20	80	1.06	1.02
30	70	1.13	1.04
40	60	1.31	1.05
50	50	1.31	1.07
60	40	1.44	1.09
70	30	1.61	1.11
80	20	1.85	1.13
90	10	2.22	1.16
91	9	2.27	1.16
92	8	2.32	1.17
93	6	2.38	1.17
94	6	2.44	1.17
95	5	2.51	1.18
96	4	2.58	1.18
97	3	2.65	1.18
98	2	2.74	1.18
99	1	2.83	1.19
100	0	2.93	1.19

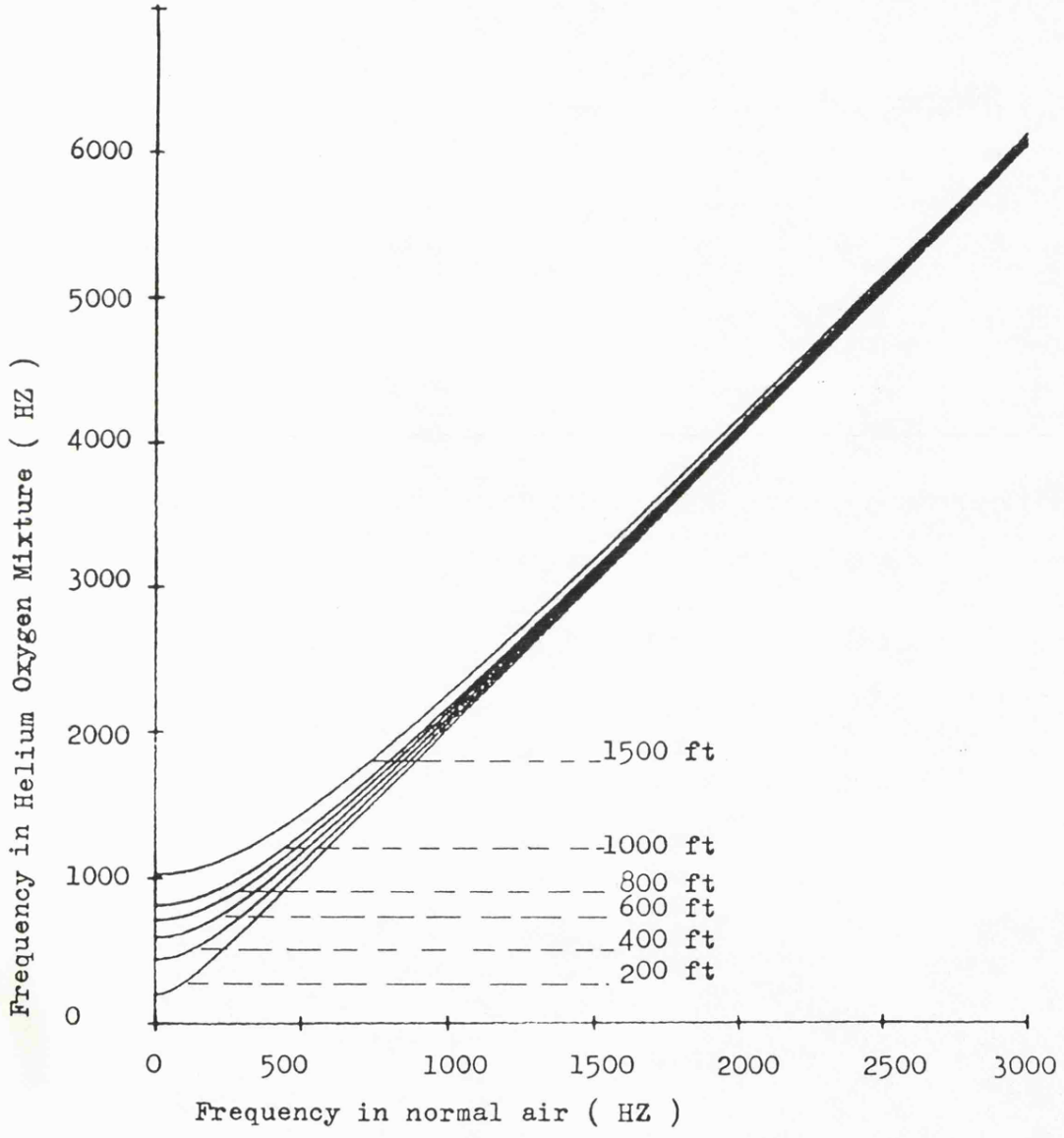


Figure 2.11 Formant Frequency Shift in a Mixture of 90% Helium ,10% Oxygen and for various depths.

2.4.2 Formant Bandwidth

In 1982 Richard[35] showed theoretically that the formant bandwidth of Helium speech increases nonlinearly in a Helium Oxygen environment. His prediction was based on the speech production model given by Flanagan[3]. He estimated that the low frequency formant bandwidths increase as much as K_1^3, while the higher ones by as much as K_1 (Figure 2.12). At the same time report by the Norwegian Underwater Technology Centre[19] based on actual measurements of the Helium speech confirmed Richard's finding. They showed an increase by a factor of 14.7 (K_1^3) in the lower formant bandwidth and by a factor of 1.2 in the higher formant bandwidths. Both reports contradicts the earlier claim by Quick[53] and Giordano et al.,[31] that there were no changes in the formant bandwidths.

The present Helium speech unscrambler compresses the envelope spectrum linearly which results in a linear reduction in the formant frequencies and bandwidths. This linear compression enhances the processed helium speech relative to unprocessed speech, but the distortion and unnaturalness associated with processed speech could be the result of improper correction in formant frequencies and bandwidths[26]. The formant bandwidth is not a critical factor in normal speech intelligibility provided they are not very large[40]. However, its effect on the Helium speech intelligibility is unknown.

2.4.3 The source transfer function:

Two factors affect the excitation sources: they are the pitch and source spectrum. A small change in the pitch of Helium speech has been observed by many researchers[17, 20, 21]. It was attributed to behavioural changes rather than to physical

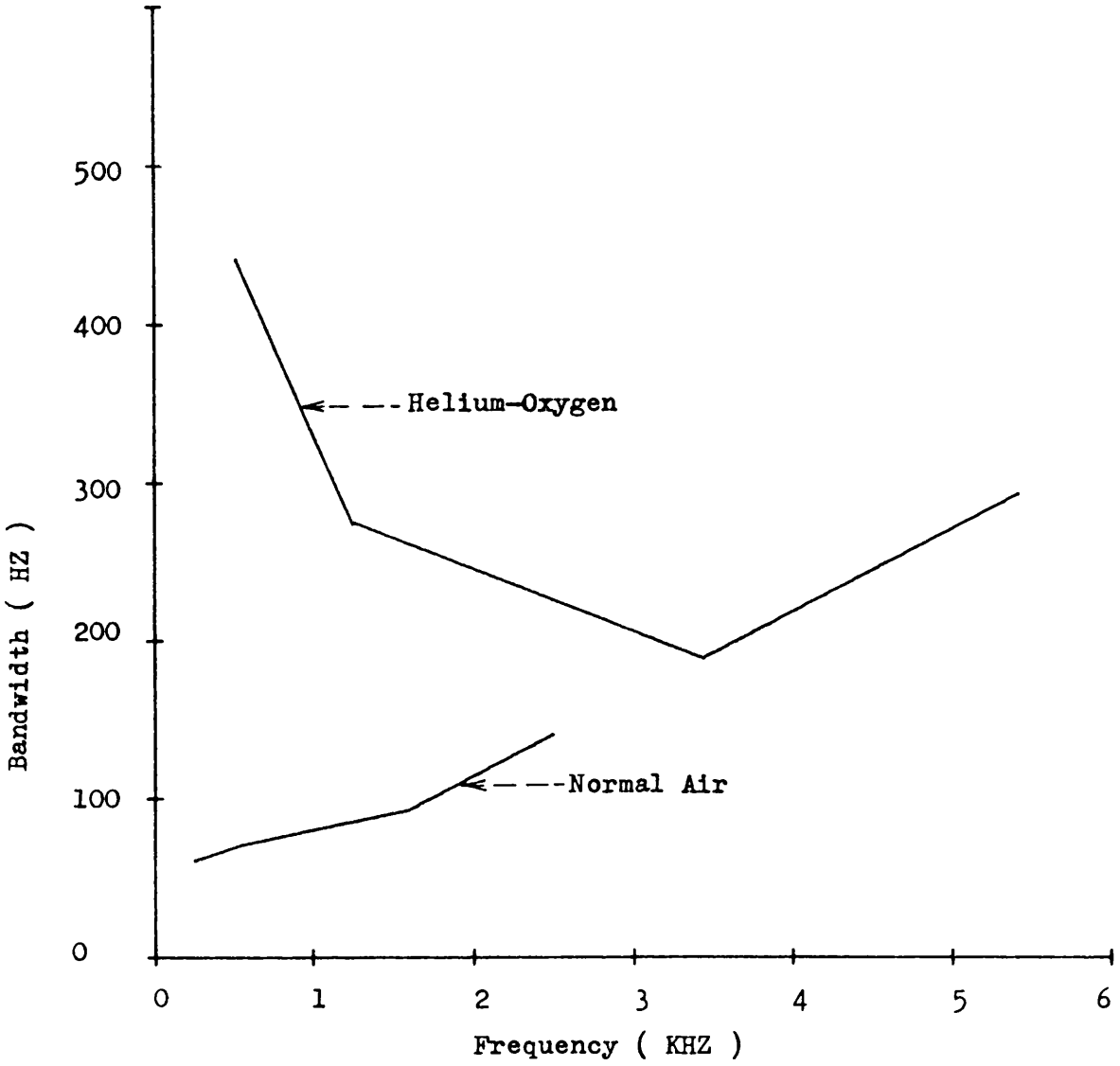


Figure 2.12 Formant Bandwidth in a Helium Oxygen and Normal Air Enviroments.
(After Richard (35))

conditions[17, 24], because it was known that a diver will tend to alter his pitch period in an attempt to enhance his speech intelligibility[38]. In fact this change in the pitch has only a slight effect on the speech intelligibility[21, 35, 38].

The only information available on the effect of Helium oxygen environments on the source spectrum is described in a study by Tanaka et al.,[24]. He concluded that this environment had no effect on the glottal pulse spectrum and hence the transfer function G(S). Inserting this information into the transfer function given by equation 2.5 and equation 2.6 shows a decrease in the level of low frequency formants relative to the high frequency formants. This result contradicts well documented observations that the high frequency level decreases relative to the low frequency one in Helium speech[17, 22].

There are two reasons that could account for this obvious anomaly. Firstly Fant and Sonesson[22] in 1964 suggested that the amptitude response of the vocal tract function to voiced sounds was proportional to $1/\sqrt{\rho}$ whilst to unvoiced sounds it was proportional to $1/\rho$ If this assumption is applied to the transfer function of the radiation given by equation 2.8, then the voiced sound change would be proportional to $\sqrt{\rho}$ whilst the unvoiced sound change would be independent of depth and therefore subject to a relative reduction by a factor of $\sqrt{\rho}$.

Secondly it could be explained by a degradation in the high frequency response of the microphone used in the divers helmet[35]. However, there may well be other explanations as yet unreported.

Therefore, it is customary to consider only empirical formula for amplitude equalisation in enhancement systems.

2.5 CONCLUSION

Speech originates with vibrations of the vocal cords or with constriction of the air flow from the lungs. These are filtered in the vocal tract and radiated through the lips. The resonant frequencies of the vocal tract are called formants. The first three formants are the most important elements affecting speech intelligibility. The vocal tract moves slowly with time to produce different sounds.

The ear resolves the speech signal into frequency bands. These bands are called critical bands. Over short time intervals the energy in these bands is important for speech perception. Whereas the phases of different frequency components are unimportant for speech intelligibilty.

Low level signals can be masked by the presence of a more intense signal or by the presence of noise. The ear is highly sensitive to tones presented to it simultaneously, whilst this sensitivity decreases with an increase in frequency.

Speech formant frequencies shift upward nonlinearly in the Helium Oxygen mixture used for deep sea diving. The bandwidths of these formants increases drastically whilst the level of the high frequency spectrum is attenuated in these environments. The Helium oxygen environment has very little effects on the speech pitch.

CHAPTER THREE

HELIUM SPEECH UNSCRAMBLERS

3.1 INTRODUCTION

In the past twenty years large numbers of Helium speech unscramblers have been reported. They are either operated in real time or have been simulated on a computer. The review of these systems enables a designer to gain some insight into the present state of the art in the design of these systems. It also assists an understanding of the effect of certain speech parameters on speech intelligibility.

Helium speech unscramblers can be classified into two broad types which operate either in the time domain or in the frequency domain. Time domain unscramblers rely on processing the speech waveform to achieve enhancement of Helium speech, whilst frequency domain unscramblers rely on processing the speech spectrum.

3.2 TIME DOMAIN TECHNIQUES

3.2.1 Recording and Play Back Technique

One of the earliest techniques developed to process Helium speech operated in the time domain[27]. The Helium speech was recorded at one speed on a magnetic tape recorder and then played back at slower speed. Although it was simple and an increase in intelligibility was reported, it cannot clearly be used in real time. It decreases the pitch in proportion to the amount of speed reduction which affects the naturalness of the speech.

3.2.2. Waveform Segmentation and Expansion

The disadvantage of the recording-play back technique was overcome later by taking only segments of Helium speech and expanding them in time by a constant factor, in synchronism with the speech pitch period[29, 33, 34, 54, 55] as shown in Figure 3.1. Systems dependant on this technique could now be easily implemented in real time using digital techniques.

The synchronous operation of real time systems preserve the pitch of the speech. However, these systems cannot correct formant frequencies non linearly and they do not possess the ability to compensate for the high frequency response. Another difficulty associated with these systems is pitch detection. If the system looses synchronisation due to incorrect location of the start of each pitch period, then the unscrambled speech could be masked in noise[38]. The use of simple pitch detectors which detect maxima in the speech waveform could cause this problem due to the increase in formant frequencies[53]. Hence extra care is needed in designing pitch detectors for a Helium speech unscrambler.

3.2.3 Auto Correlation Techniques

In early time domain systems the expansion was performed on the time waveform directly. However, an alternative technique, simulated on a computer, was used to process the autocorrelation of segments of the Helium-speech[32]. This autocorrelation, which consists of the same frequency components as the Helium speech waveform is expanded by a method similar to the previous one. This technique suffers from the same disadvantages as the waveform segmentation and expansion technique, however it can reduce the

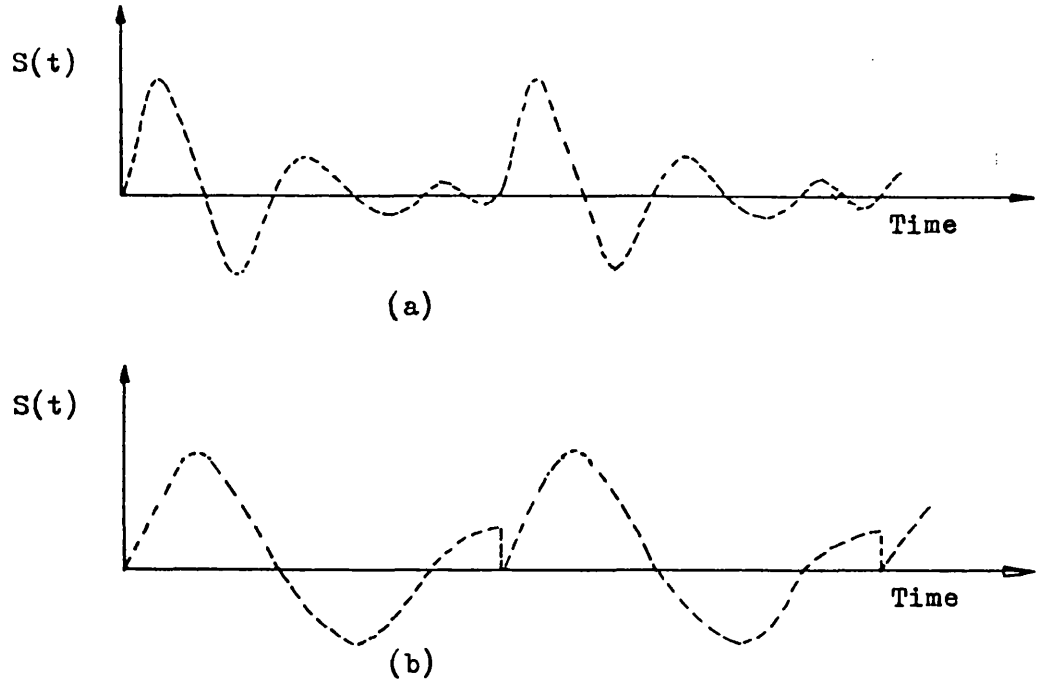


Figure 3.1 Principle of Time Domain Unscramblers.
(a) Signal in Helium - Oxygen
(b) Unscrambled Signal

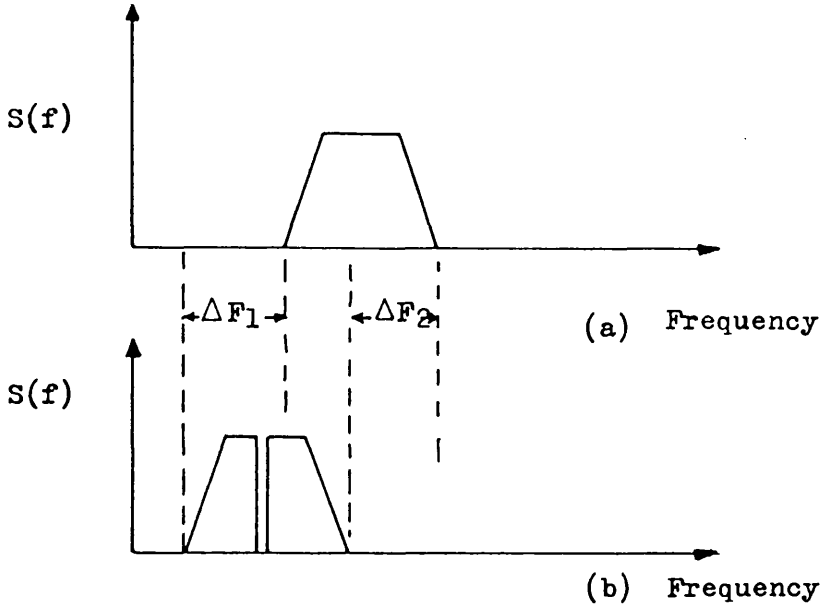


Figure 3.2 Principle of Frequency Subtraction Unscrambler.
(a) Helium Speech Spectrum
(b) Unscrambled Speech Spectrum

output noise level by discarding selected portions of the autocorrelation function. It also has the ability to modify the speech pitch which can be useful if the pitch is interferring with the intelligibility of the speech.

3.3 FREQUENCY SUBTRACTION TECHNIQUES

In the frequency subtraction technique the entire Helium speech spectrum is moved down in frequency by a controllable manner and filtering process.

A modified and improved version of this technique[28] is shown in Figure 3.2. In this method the Helium speech spectrum is separated into sub-bands prior to processing. Then these sub-bands are moved down in frequencies separately.

These systems are the simplest implementation of the frequency domain processing techniques. They shift formant frequencies downward rather than dividing it, and hence preserving the formant bandwidth. This will have a greater effect on speech intelligibility if the formant bandwidths are large. Also the technique alters the pitch of the unscrambled speech and produces a rather poor performance.

3.4 Vocoder Techniques

The voice coder or Vocoder has been used extensively in the past as the principal bandwidth compression device, to achieve low bit rate speech transmission and to produce speech synthesisers. There are several different types of vocoder all of which have the ability to extract the envelope of the speech signal, determine its

excitation sources separately and hence provide the facility to modify the envelope structure without altering the excitation sources. This facility had been used in at least three Helium speech enhancement systems. These are the Hustle, the FRV and the Voice Transcoder.

3.4.1 Hustle

The Helium underwater speech translating equipment[30] is a modified version of the conventional channel vocoder used for normal speech processing. As shown in Figure 3.3 the Helium speech is passed through a bank of bandpass filters (analysing filters) with bandwidth and centre frequencies K time greater than that of the synthesiser's filters. The output of these filters is applied to the envelope detectors which produce outputs proportional to the energy in each frequency band. These signals are then used to modulate a source derived from the Helium speech signal by the pitch detector. The pitch detector determines whether the input speech is voiced or unvoiced. For the voiced speech a periodic signal is generated, while the unvoiced speech is represented by a random noise generator. The modified signal is passed through synthesising filters with their centre frequencies and bandwidths compressed by a factor K. The resulting signals are recombined to produce the enhanced speech.

This technique was reported by Rowerth, 1969 and it was operated in real time. It used twenty two bandpass filters on each side and a compression ratio of K equal to 2.2.

Generally this type of unscramblers are only able to correct linear distortion. They are constructed to correct for a single

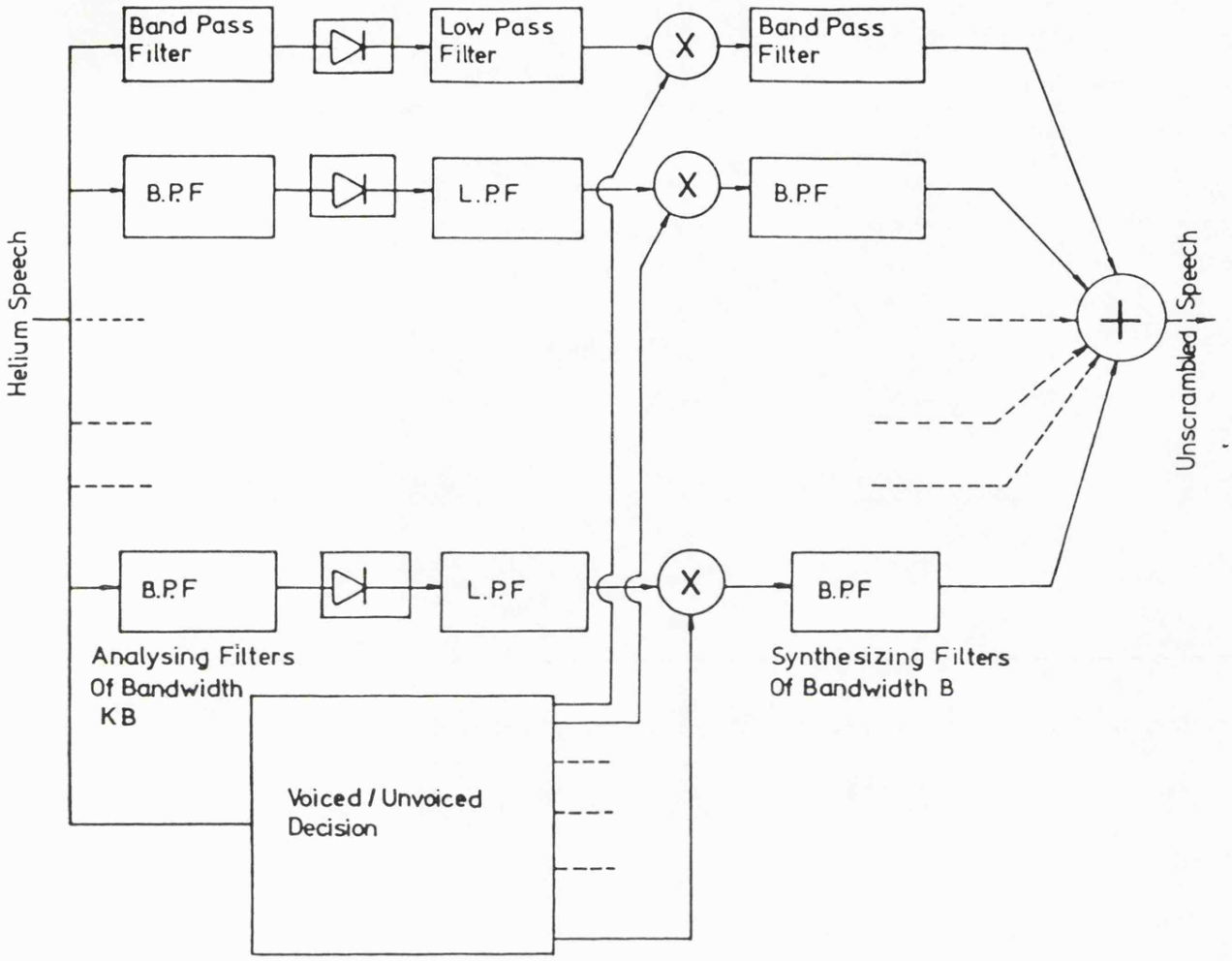


Figure 3.3 Helium Speech unscambler using conventional vocoder (Hustle).

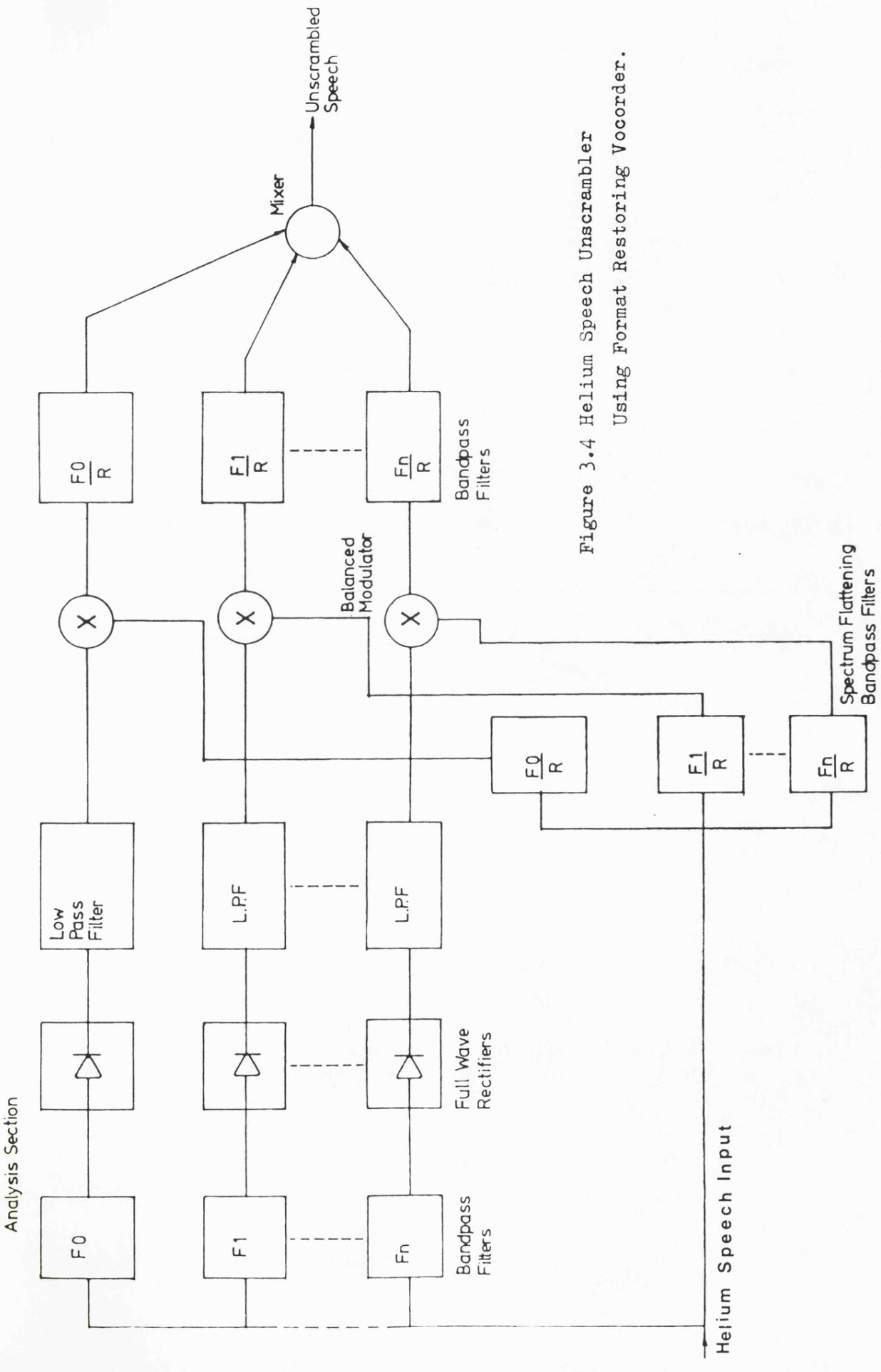


Figure 3.4 Helium Speech Unscrambler
Using Format Restoring Vocorder.

value of helium oxygen mixture and hence diving depth. The difficulty of measuring pitch and determining the correct excitation sources in the noisy diving environment leads to a significant loss of intelligibility.

3.4.2 Formant Restoring Vocoder (FRV)

The disadvantages of measuring pitch and making voiced discrimination in Hustle had been avoided in this enhancement system. The FRV uses an identical technique to that used in Hustle except that the modulated signals deriving the synthesising filters are the product of the energy in each channel and a signal derived from the input speech using a spectrum flattening circuit as shown in Figure 3.4. This spectrum flattening circuit generates a signal with equal amplitude harmonics. These harmonically related signals are separated by filters identical to those on the analyser side of the FRV.

An enhancement system based on this technique has been reported by Golden(1966)[56]. The system is simulated on a computer. Again, this unscrambler is useful only for a specific Helium oxygen mixture and again cannot correct the non linear frequency distortion.

3.4.3. Voice Transcoder

An interesting system which is based on the voice excited vocoder had been proposed by Zucher[57](1980). This system, unlike other vocoders, has the ability to select six different compression ratios.

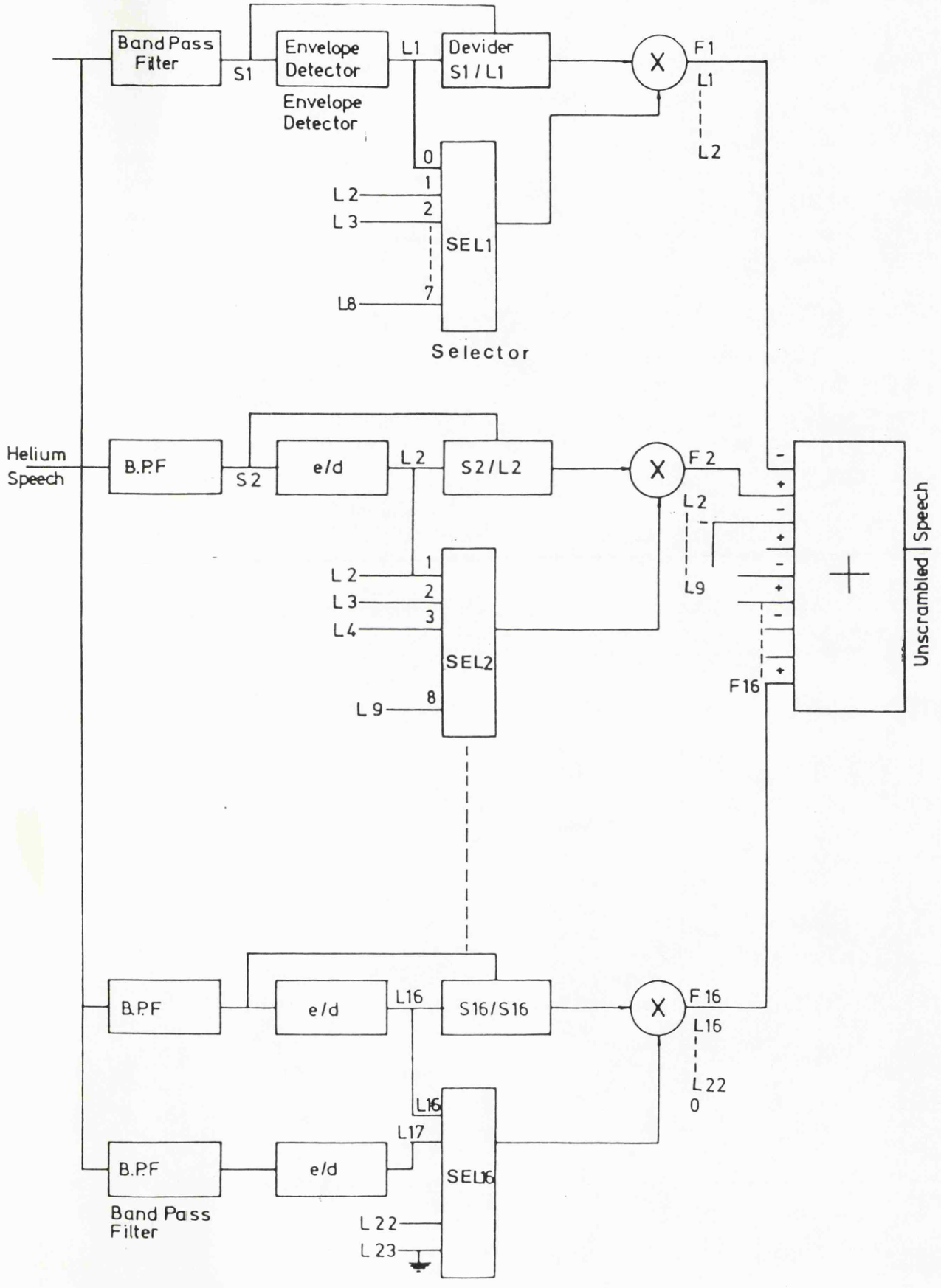


Figure 3.5 Helium Speech Unscrambler Using Voice Transcoder.

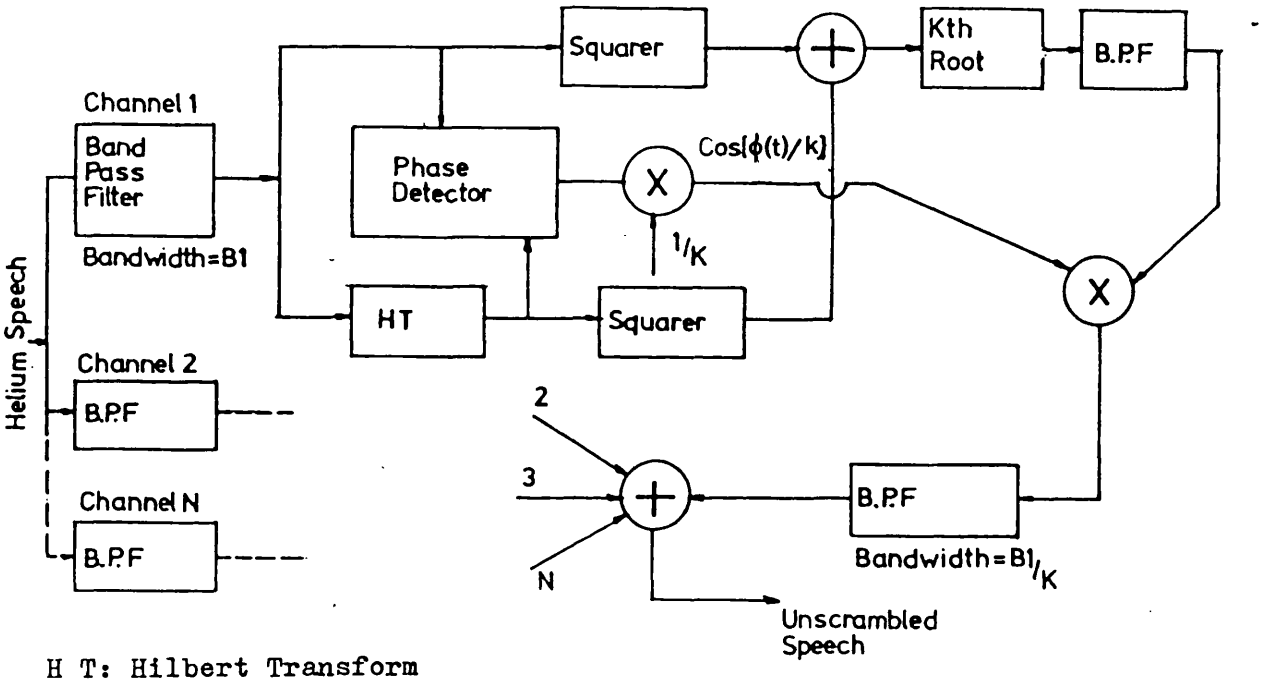
The system as shown in Figure 3.5 uses sixteen main channels and six auxillary channels. It divides the maximum possible expansion factor in the Helium speech spectrum into six discrete values. It assumes that the relation between centre frequencies of the adjacent filters is equal to this value. To compress the speech spectrum the measured envelopes at different channels are redistributed. For example if the Helium speech is expanded by the expansion factor number two then it will be assumed that the envelope of channel two in the helium oxygen mixture belongsto the envelope of channel one in normal conditions, while the envelope of channel three belongs to channel two etc. The frequencies for each channel are derived from Helium speech by a method similar to that of the FRV. The output of each main bandpass filter is divided by its envelope to obtain an appropriate equal amplitude signal.

However, this system again only corrects linear distortion, although it could be made to compensate for amplitude attenuation by inserting amplifiers in the envelope detector circuits. The use of only one bank of filters greatly reduces the hardware required to build the transcoder.

3.5 ANALYTICAL SIGNAL ROOTING TECHNIQUE

Analytical signal rooting is a technique by which the speech signal is split into a number of frequency bands with the intention that at at any time only one formant will be present in each band. then the signals in these bands are amplitude rooted as a means of compressing their spectra.

The analytical rooting operation can be explained by defining the signal S(t) in each band by:



H T: Hilbert Transform

Figure 3.6 Helium Speech Unscrambler Based on Analytical Signal Rooting.

$$S(t) = A(t).e^{j\phi(t)} \qquad 3.1$$

Taking the nth root of equation 3.1 would achieve an approximate reduction in the bandwidth of the signal by factor n. Thus,

$$[S_n(t)]^{\frac{1}{n}} = [A(t)]^{\frac{1}{n}} . e^{j\frac{\phi(t)}{n}} \qquad 3.2$$

This method has been suggested by (Flanagan, Griondo, Flanagan and Bogner , Mahala)[3, 31, 14, 58] for Helium speech enhancement systems. An enhancement system based on this principle was simulated on a computer [59] as is shown in Figure 3.6. In this system band pass filters are used to separate the Helium speech into three bands. Then Hilbert transformation is used to separate A(t) and $\phi(t)$ in each band. The nth root of both amplitude and phase is taken and then the modified signals are recombined to produce the enhanced speech. In simulated systems, only linear transformation of formant frequencies has been attempted without amplitude correction. However, it was claimed that this system improved speech intelligibility. It is interesting to note that, due to the expansion of the Helium speech bandwidth, the pitch appears distinctively as the envelope of the output from the bandpass filter. The rooting of the envelope will have very little effect on the speech pitch, because the time variation of the envelope will be preserved. This phenomena was observed by Takasugi et al., [59] and reported by Falangen, 1974, Falangan and Bonger, [3, 14].

3.6 FAST FOURIER TRANSFER TECHNIQUE

The fast fourier transform (FFT) is an efficient algorithm by which the fourier transform of a signal can be computed. It is finding increasing application in digital signal processing due to its efficiency in reducing computation time required for calculating the fourier transforms. However, determining the freqency components of a complex signal using the FFT technique has clearly more flexibility than a technique that relies on subjection of the signal to a bank of bandpass filters and examining the output of each. Although overall the processes are identical. Computer programs that can be used to determine these parameters form an efficient tool for investigating speech characteristics generally. However, the main disadvantage of these techniques is the considerable computation time required to calculate FFT parameters.

However, a Helium speech enhancement system has been proposed recently by Richard (1982)[35, 18] and implemented in real time by Norwegian Underwater Technology Centre[19]. It is the first reported attempt at correcting many of the known forms of Helium speech distortions simultaneously. The technique is used to determine the envelope of the speech spectrum by calculating the spectrum. Then an algorithm is used to compress the calculated spectral values non-linearly, and modify the amplitude appropriately prior to reconstruction in the normal speech bandwidth. However, work is yet to be done to optimise this system as the translated speech is unclear in places and lacks of crispness [19].

Apparently no conclusions have been drawn about the effectiveness of non linear correction of formant frequencies. It is, however, the only system that includes non linear correction in

the original design. This system also has the ability to correct formant bandwidth and reduce Helium speech noise. It is interesting to note that the intelligibility was reduced when a noise reduction algorithm was used to subtract noise from the Helium speech.

3.7 CONVOLUTION TECHNIQUE

In this technique the envelope spectrum of the Helium speech is separated from the excitation sources by a method known as Homomorphic Technique. Then the envelope spectrum is modified and convolved with the excitation sources to produce enhanced speech.

The principle behind this technique is shown in Figure 3.7. The spectrum of a short segment of Helium speech is calculated by a FFT. The resulting spectrum is the product of the fourier transform of the excitation and the vocal tract impulse response. The logarithmic magnitude of the spectrum is then computed which converts the multiplication of excitation spectrum and vocal tract spectrum into a summation. The inverse fourier transform of the logarithmic magnitude produces a new function which is called cepstrum. The cepstrum consists of two components, the slowly varying component which represent the slow movement of the vocal tract and a rapidly varying component which represents the excitation sources.

The vocal tract related components of the cepstrum are separated from the excitation component by appropriate filtering. Then the low frequency components are fourier transformed and subjected to exponential characteristics to produce the vocal tract

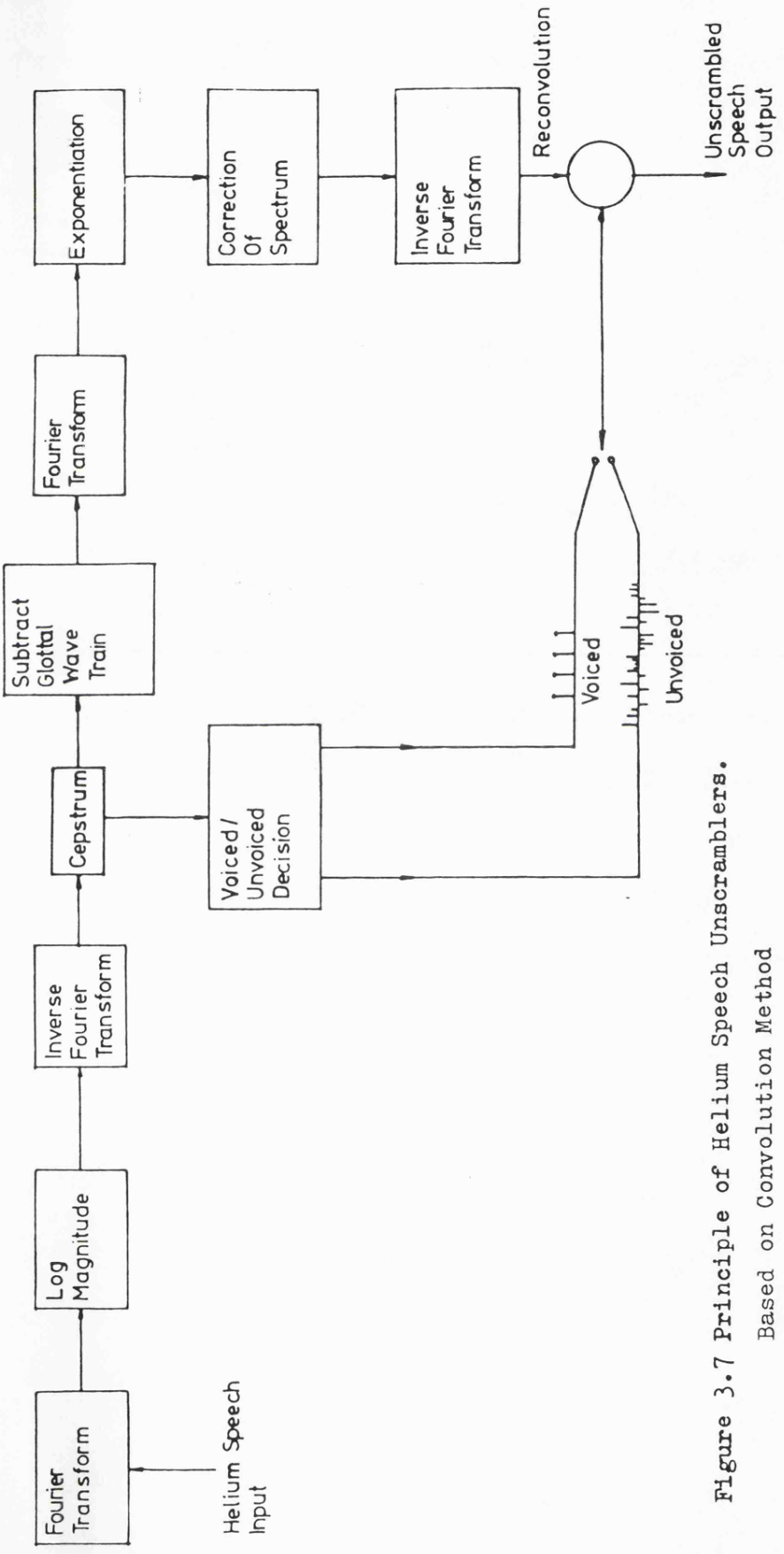


Figure 3.7 Principle of Helium Speech Unscramblers.

Based on Convolution Method

transfer function. This function is then modified to produce the impulse response of the vocal tract.

The pitch and voice/unvoiced decision can also be derived from the cepstrum. In voiced sounds the distance between the amplitude peaks is an indication of the pitch period, whilst the absence of these peaks is an indication of unvoiced sounds.

Finally convolution of the modified impulse response of the vocal tract with appropriate excitation produces the enhanced speech.

Clearly this technique is highly complex and would require considerable processing hardware. However, an off-line system based on this technique has been simulated on a computer[53]. The simulated part corrects the envelope non linearly. While the pitch information had been entered manually.

3.8 LINEAR PREDICTIVE TECHNIQUE

In this technique a short segment of Helium speech is passed through a filter the frequency response of which is the inverse of the envelope of the speech spectrum. The output from this filter is a signal which contains the source information. The filter parameters are predicted for each new segment of the speech signal.

The block diagram of this technique is shown in Figure 3.8. The envelope spectrum at A is subsequently modified and this corrected spectrum at C is used to calculate the coefficients of a synthesiser filter representing the envelope spectrum of the enhanced speech signal. This enhanced speech which is determined by the excitation source is produced at the output of the controlled synthesiser filter. Although this technique has the ability to

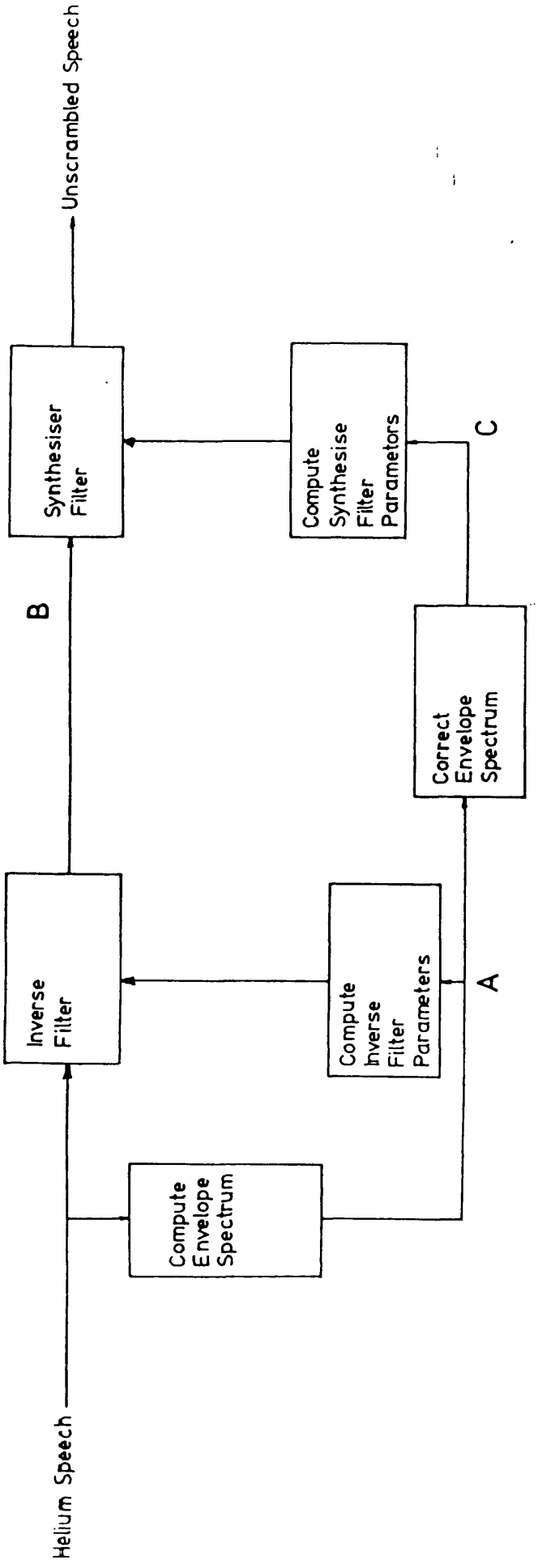


Figure 3.8 Principle of Helium Speech Unscrambler
Based on Linear Prediction Technique.

correct non linear frequency distortion[60], enhancement systems based on this technique have only been used to correct linear distortion. These systems were simulated on a computer and would not appear to be feasible for real time implementation[39].

3.9 CONCLUSION

There are two broad types of Helium speech unscramblers. Time domain unscramblers which are generally more suitable for real time implementation inherently have no facility to correct non linear distortions of the formant frequencies or formant bandwidths. Frequency domain unscramblers process the Helium speech spectrum. They are generally far more complex than the time domain unscramblers. However certain of them have the potential to equalise non linear distortions associated with Helium speech[35, 37, 53].

Although current real time unscramblers do not achieve non linear frequency compression, they enhance Helium speech relative to unprocessed speech. However, the processed Helium speech remains unnatural with degraded intelligibility.

CHAPTER FOUR

THE THEORY OF SPEECH COMPRESSION

4.1 INTRODUCTION

Helium speech as described in Chapter two suffers from a unique distortion of its formants. This distortion affects the frequency domain parameters of the speech. The envelope spectrum of the speech expands non linearly and occupies a different band in the frequency domain. Whereas the fine structure of the envelope spectrum (i.e. the harmonics) retain the same relationship to that of normal speech. Thus any technique which aims to equalise all of the speech distortions should ideally be based on a frequency domain technique. Also it should have the ability to modify independently different frequency components of the Helium speech signal.

The technique which processes the speech in the frequency domain is called the analysis-synthesis technique. Analysis-synthesis produces an accurate representation of the speech production mechanism. It divides the speech signal into several frequency bands with a bank of band pass filters. The signals in these bands are then described by a set of parameters which reflect the physical characteristics of the speech production mechanism.

As a helium speech processing technique, analysis-synthesis is complex compared with other techniques such as time domain techniques and requires complicated circuitry to be implemented in real time. However it is the only technique with the capability of equalising the non-linear Helium speech distortions[33, 34, 64].

An interesing characteristic associated with certain analysis-synthesis systems is their ability to manipulate the parameters of

the speech signal[1]. This is potentially useful in unscramblers which use software rather than hardware to implement the correction algorithm. This feature is very useful in real time Helium speech unscrambling because different diving conditions require different correction algorithms. Clearly, algorithms based on software are easier to modify than those based on hardware. Therefore such systems could also be used to determine the contribution of different distortions on the Helium speech intelligibility.

4.2 PRINCIPLE OF ANALYSIS-SYNTHESIS

The representation of the speech signal by a sinusoidal waveforms has been used widely[3, 62] as a mathematical tool for defining the different parameters associated with speech processing techniques. The method of this representation is useful in linear systems, such as speech, because it is possible to resolve the signal into its component parts which can be processed separately prior to reconstruction by addition. This procedure, which is known as superposition, is useful in practical systems because it can be related to the physical properties of the speech production mechanism. If for example the signal is represented by:

$$S(t) = \sum_k A_k(t) e^{j\phi_k(t)} \qquad 4.1$$

then $A_K(t)$ and $\phi_k(t)$ are chosen to reflect the physical properties of the mechanism. Therefore the changes in $A_k(t)$ and $\phi_k(t)$ are effectively the same as the changes in the physical properties of

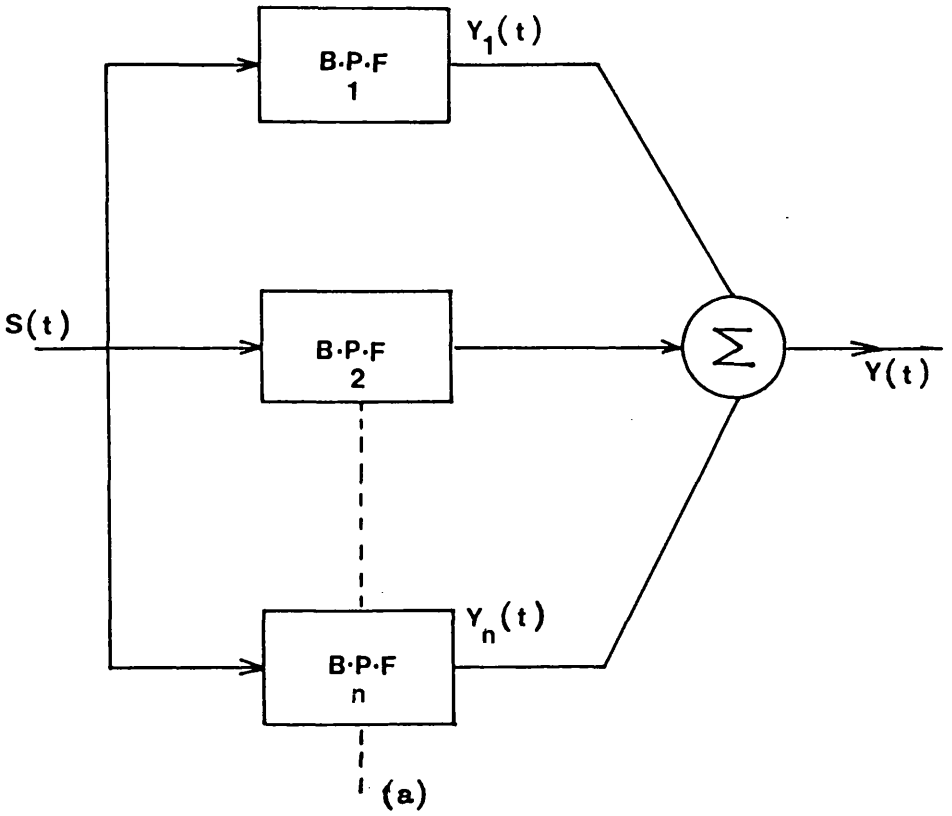
speech. Conversely changes in the physical properties of the speech can be determind by detecting changes in $A_K(t)$ and $\phi_K(t)$.

In 1966, Flanagan and Golden[15] in proposing a frequency compressing technique, were the first to establish mathematically a relationship between the physical changes in the speech production organ, and the speech signal represented by equation 4.1. In his specificaton for the "phase vocoder", he proposed that the bandpass filters shown in Figure 4.1 were distortionless. Then the output Y(t) would be a good approximation to the input speech. To satisfy this distortionless condition the overall response must provide no more than ±2db ripple over the speech band and constant group delay characteristics.

In fact phase vocoder theory became the basis of many other systems in the speech processing field[12, 45]. It has also been used to describe earlier systems such as the vocoder[4], Vobac[10] and the Codimax[11].

4.2.1 SHORT TIME FOURIER TRANSFORM

As described in Chapter Two, the speech waveform is the response of a linearly time varying system. This system changes slowly with time, therefore the structure of the speech waveform can be considered stationary over a short time interval. The output of this system can be represented by a conventional fourier transform over this short interval. This representation is called the short time fourier transform (STFT). However, the fourier transform is a linear operation and hence the fourier transform of the speech signal at the output of each filter of Figure 4.1 represents individual parts of the fourier transform of the speech signal.



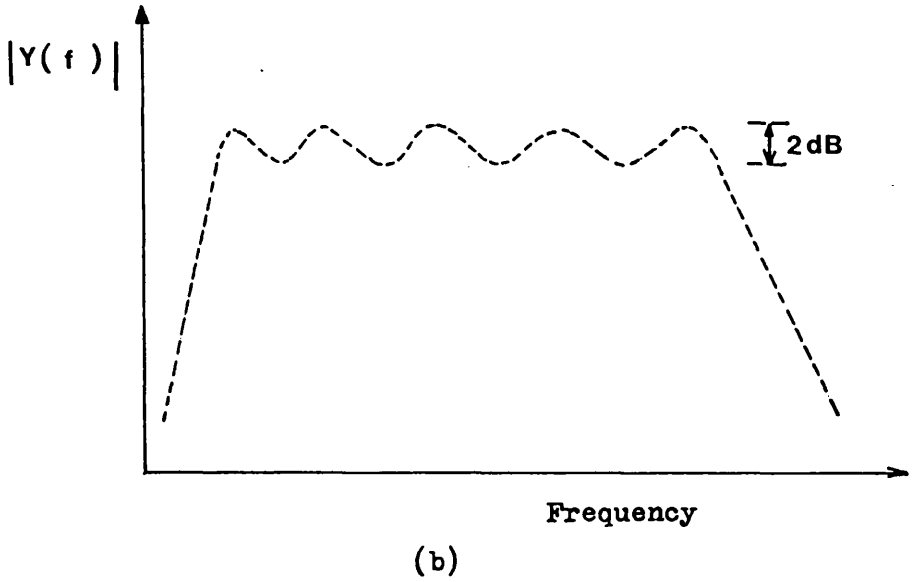


Figure 4·1 Analysis of Speech

(a) Contiguous Band Pass Filter

(b) Overall Frequency Response

The output Yn(t) of each filter is the convolution of the impulse response of that filter with the input speech signal. Mathematically this output is given by:

$$Y_n(t) = \int_{-\infty}^{t} S(\tau) \cdot h(t-\tau) \cdot d\tau \qquad 4.2$$

where $Y_n(t)$ is the output of the nth filter and $h_n(t)$ is the impulse response of that filter. If the bandpass filters have identical bandwidths and phase characteristics then their impulse responses can be given by:

$$h_n(t) = h(t) \cdot \cos w_n t \qquad 4.3$$

where $h_n(t)$ is the impulse response of a low-pass filter and w_n is the centre frequency of the nth bandpass filter. Substituting 4.3 into 4.2 gives:

$$Y_n(t) = \int_{-\infty}^{\infty} S(\tau) \cdot h(t-\tau) \cdot \mathrm{Cos}(w_n(t-\tau))\, d\tau \qquad 4.4$$

or

$$Y_n(t) = R_e\left[e^{jw_n t} \int_{-\infty}^{\infty} S(\tau) \cdot h(t-\tau) \cdot e^{-jw_n \tau} d\tau\right] \qquad 4.5$$

The integral in equation 4.5 is the fourier transform of the part of

the signal in the window h(t), and evaluated at a frequency w_n[15] Denoting this term by $X_n(w_n, t)$ and rewriting equation 4.5 gives:

$$Y_n(t) = Re\left[X_n(w_n,t)\ e^{jw_n t}\right] \qquad 4.6$$

$X_n(w_n,t)$ is the short time amplitude and phase of the speech spectrum, and can be represented by :

$$X_n(w_n,t) = |X_n(W_n,t)\ |e^{j\theta_n(w_n,t)} \qquad 4.7$$

Introducing this new representation of $X_n(w_n,t)$ into equation 4.6 gives:

$$Y_n(t) = |X_n(w_n,t)| \cdot Cos\left[w_n t + \theta_n(w_n,t)\right] \qquad 4.8$$

Equation 4.8 is the basic equation for all analysis synthesis techniques. The way by which the STFT is derived from the speech signal defines the processing technique. Also the STFT reflects the characteristic of the speech production mechanism. The usefulness of equation 4.8 can be demonstrated by assuming that narrow band filters are closely spaced and that the speech pitch is constant. If only one harmonic is present at the output of each filter, then $|X_n(w_n,t)|$ reflects the slowly varying amplitude response of the vocal tract at frequencies approximately equal to w_n's. The phase derivative would be constant and represent the deviation of the frequency of harmonic components from the centre frequency. The vocal tract and pitch are varying slowly in normal speech. Then both the amplitude and phase will vary slowly also. The amplitude

now reflects the slow variation of the vocal tract, whilst the phase derivative reflects the slow variation of the excitation source or the fine resolution of the speech spectrum i.e. harmonics.

For wide band filters which pass more than one harmonic, the parameters will reflect equally the slow variation of both the vocal tract and the excitation characteristics.

Finally equation 4.8 can be used to describe a generalised configuration for the analysis-synthesis systems. In the analyser section the amplitude and phase spectra of the speech signal are determined from the outputs of the bank of band pass filters. These parameters are then used in the synthesiser section to modulate the phase and amplitude of a set of sinusoidal generators. The outputs of the modulated sources are then summed to obtain the synthesised speech as shown in Figure 4.2.

Processing techniques depends on modifying the amplitude and phase of the short time fourier transform prior to synthesis. If for example narrow band pass filters are used in the analyser the envelope spectrum of the speech signal can be scaled by modifying only the amplitude of the STFT[35] whilst if wide band filters are used to analyse the speech, both the amplitude and phase of the STFT must be modified to scale the envelope spectrum of the speech signal[12]

The magnitude of the STFT can be measured by a simple envelope detector in each band[3] whilst the phase, which is necessarily a relative parameter cannot be defined absolutely. In most speech processing systems the phase is calculated from a measurement of the instantaneous frequency in the appropriate band. This calculation

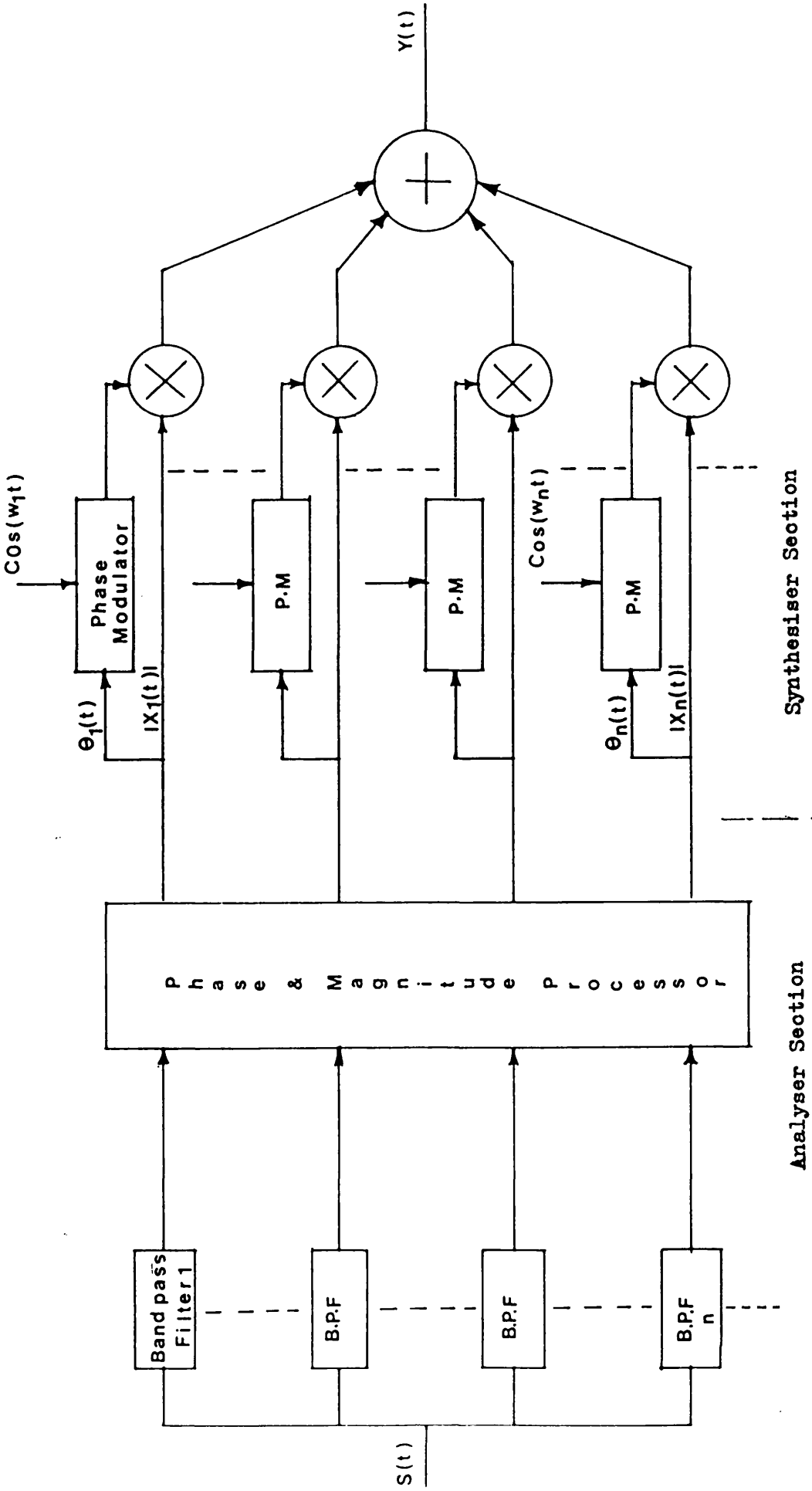


Figure 4.2 Generalised Configuration for an Analysis - Synthesis Processing System.

of the phase is then straight forward because integrating the instantaneous frequency gives a value of phase. The integration operation can ignore the constant associated with it without affecting speech intelligibilty due to the insensitivity of the ear to phase.

4.3 AN ANALYSIS-SYNTHESIS TECHNIQUE FOR HELIUM SPEECH PROCESSING

In Chapter 2 it was shown that the characteristics of the speech production mechanism changed in Helium Oxygen environments. Also in Section 4.2 it was demonstrated that the characteristics of this model are related to the STFT of the speech signal, specifically the characteristics of the amplitude and phase of this transform.

Based on the available information on Helium speech a relationship will be developed between the STFT of the Helium speech and that of normal speech.

The basis for developing this relationship is the assumption that the expansion in the vocal tract transfer function is the most significant element effecting Helium speech intelligibility. Then to relate the transfer functions of both the source and radiation characteristics to this transfer function to determine the overall transfer function of the Helium speech production mechanism. Finally this will be related to the speech production mechanism of normal speech.

4.3.1 Relationship between Vocal Tract Transfer Function of Helium and Normal speech

The vocal tract transfer function in a Helium Oxygen mixture can be determined by assuming that:

$$S_h = \alpha S_a \qquad 4.9$$

where s_h and s_a are the complex frequencies of the vocal tract transfer functions in both the Helium oxygen and normal air environments, whilst α is the non linear expansion factor of the vocal tract frequencies in a Helium oxygen environment.

To define α it is necessary to rewrite equation 2.16 (which relates formant frequencies in both Helium and normal speech) as:

$$F_h^2 = K_1^2 F_a^2 + F_o^2 \qquad 4.10$$

were F_o^2 is given by:

$$F_o^2 = (K_2 - 1)\cdot F_{sa}^2$$

Now α can be defined as the ratio of formant frequencies in Helium speech (F_h) to those of normal speech F_a, or:

$$\alpha = \sqrt{K_1^2 + \left(\frac{F_o}{F_a}\right)^2} \qquad 4.12$$

this expansion factor is plotted (Figure 4.3) as a function of formant frequency in normal air and for a certain Helium oxygen mixture and different depths.

In defining the complex frequencies and expansion factor of the vocal tract for Helium speech as given by equations 4.10 and 4.12 it

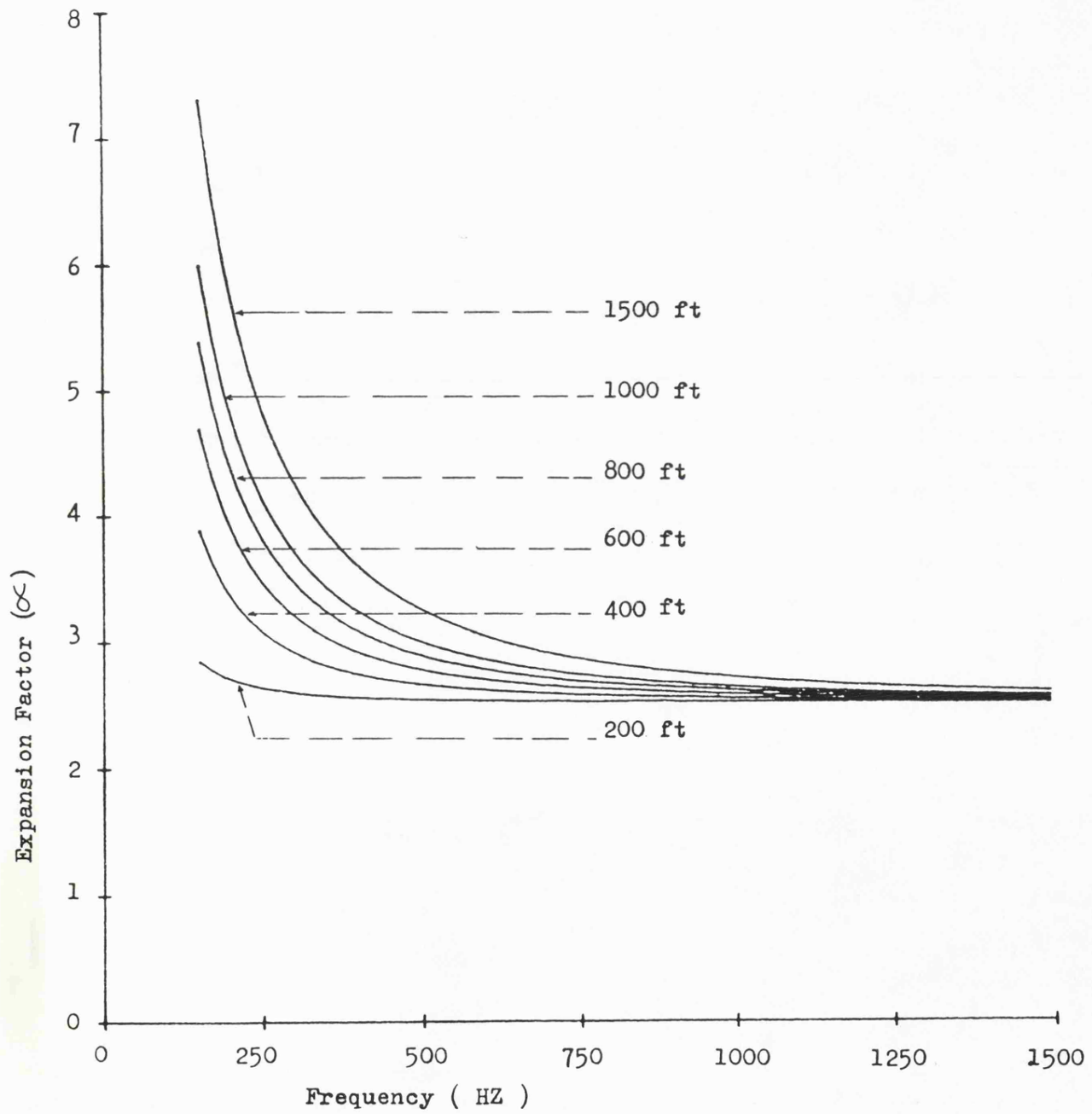


Figure 4.3 Formant Frequency Expansion Factor as a Function of Formant Frequency in Normal Atmosphere for 95% Helium, 5% Oxygen and different Depths.

was assumed that the formant bandwidths expands, by the same factor as the formant frequencies. The basis of this assumption is that, the formant bandwidth characteristics in the Helium oxygen mixture are not thoroughly understood[35]. Also all reported enhancement systems compress formant bandwidths by the speed ratio (K_1) although recent studies show a large increase in formant bandwidth at low frequency[19, 35]. It seems reasonable from the above consideration that formant bandwidth compression by a factor greater than or equal to K_1 may be adequate.

The third factor affecting the vocal tract transfer function is the formant amplitude. As described in Chapter Two empirical formula must be used to equalise this attenuation. Hence it will be assumed for the present that there is no changes in the formant amplitude of the vocal tract. Thus $V_h(s)$ can be given by:

$$V_h(\alpha S) = V_a(\bar{s}) \qquad 4.13$$

4.3.2 Relationship Between the Radiation Characteristics of Helium Speech and Normal Speech

The radiation transfer function is described by equation 2.8 as:

$$R_h(W) = \frac{0.31.\rho_h c_h}{1.44\, c_h + 1.07s} \qquad 4.14$$

In this equation the second term of denominator has only a minor

influence on the radiation characteristic, especially at low frequencies when its value is much less than the speed of the sound in the helium oxygen environment. Assuming that the complex frequency S in equation 4.14 expands by the speed ratio (K_1) in the helium environment then only a small error will be introduced into the value of the radiation characteristic. The radiation characteristics in both environments will be related by[35]:

$$R_h(\alpha s) = \frac{\rho_h \cdot c_h}{\rho_a \cdot c_a} \cdot R_a(s) \qquad 4.15$$

Equation 4.15 shows that the transfer function in the Helium oxygen mixture is increased by the same factor as the vocal tract transfer function, whilst its amplitude is increased by a constant factor dependent on the helium oxygen ratio and diving depth.

4.3.3 Relationship Between the Source Transfer Function of Helium and Normal Speech

In Chapter two it was shown that no changes in source characteristics have been reported.

Using equation 2.6 to describe the transfer function of the voiced sounds in the Helium oxygen mixture gives:

$$G_h(\alpha s) = \frac{G_a(s)}{\alpha^2} \qquad 4.16$$

also rewriting equation 2.7, to describe the transfer function of the unvoiced sounds gives:

$$G_h(\alpha s) = 1 \qquad 4.17$$

4.3.4 Relationship Between the Transfer Function of the Speech Production Model of Helium Speech and Normal Speech

The transfer function of the speech production model of Helium speech is the product of $V_h(\alpha s)$, $R_h(\alpha s)$ and $G_h(\alpha s)$. Using equations 4.13, 4.15, 4.16 and 4.17 the transfer function of the Helium speech model is given by:

$$T_h(\alpha s) = \frac{\rho_h c_h}{\rho_a c_a \alpha^2} \left[V_a(s) \cdot R_a(s) G_a(s)\right] \qquad \text{voiced} \qquad 4.18$$

and

$$T_h(\alpha s) = \frac{\rho_h c_h}{\rho_a c_a} \left[V_a(s) \cdot R_a(s)\right] \qquad \text{unvoiced} \qquad 4.19$$

The terms in brackets describes the transfer function of the normal speech production model, $T_a(S)$, thus equations 4.18 and 4.19

can be rewritten as:

$$T_h(\alpha s) = \frac{\rho_h \cdot c_h}{\rho_a \cdot c_a \cdot \alpha^2} T_a(s) \qquad \text{voiced} \qquad 4.20$$

and

$$T_h(\alpha s) = \frac{\rho_h \cdot c_h}{\rho_a \cdot c_a} \cdot T_a(s) \qquad \text{unvoiced} \qquad 4.21$$

These equations show that the transfer function of the Helium speech production model for voiced sounds is attenuated relative to the unvoiced transfer function. Although this relation is based on a well developed speech production model, the amount by which the transfer level of the voiced sounds changes is in contradiction with that observed experimentally.

The level of low frequencies are reduced relative to the high frequencies as shown in Figure 4.4 This result is in contradiction with the well known phenomena associated with observed Helium speech when the high frequencies are attenuated more than the low frequencies. As explained in Chapter two this contradiction could be the result of factors yet to be discovered affecting either the vocal transfer function or the source transfer function. However, empirical formula need to be introduced into the Helium speech transfer function to equalise this attenuation.

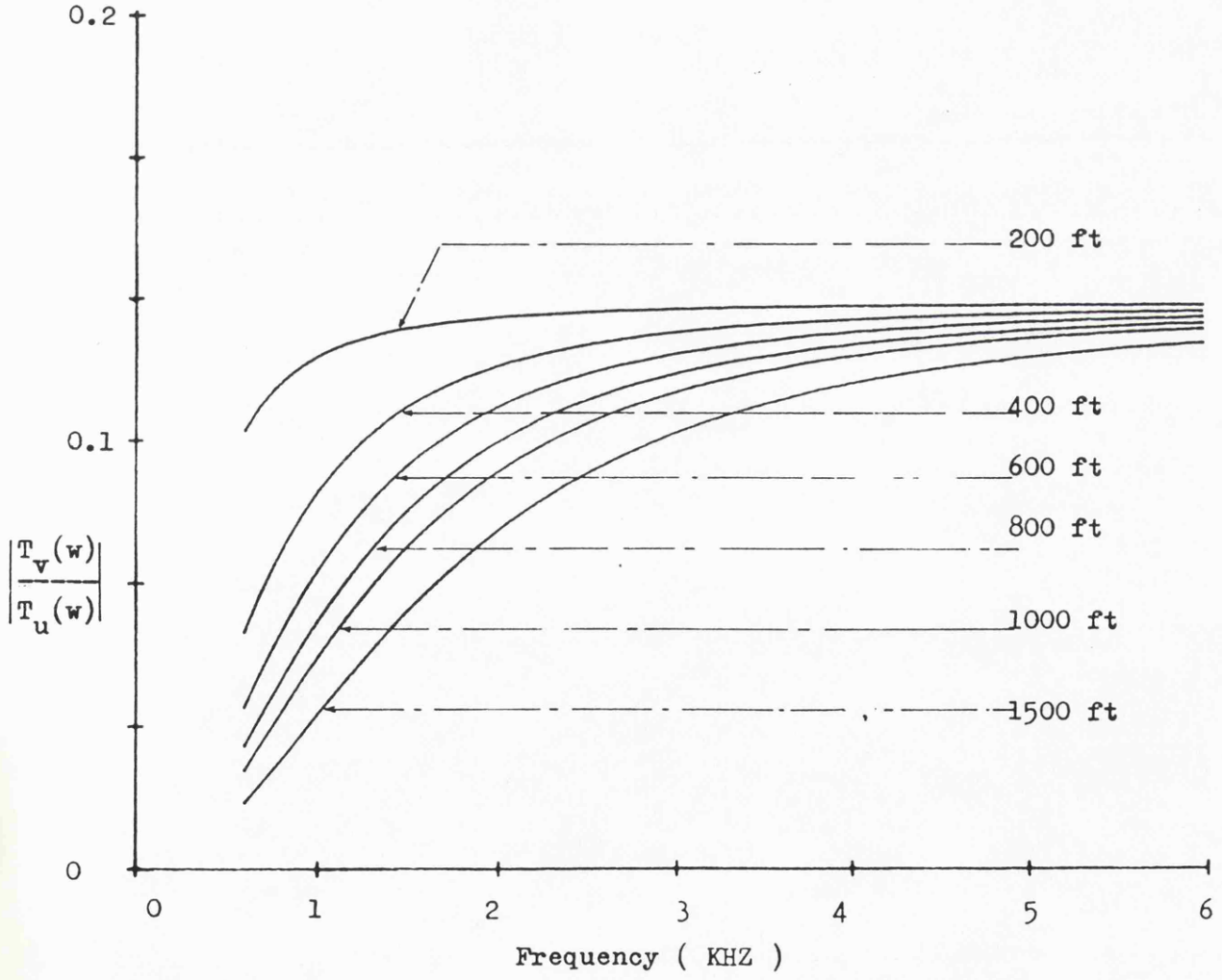


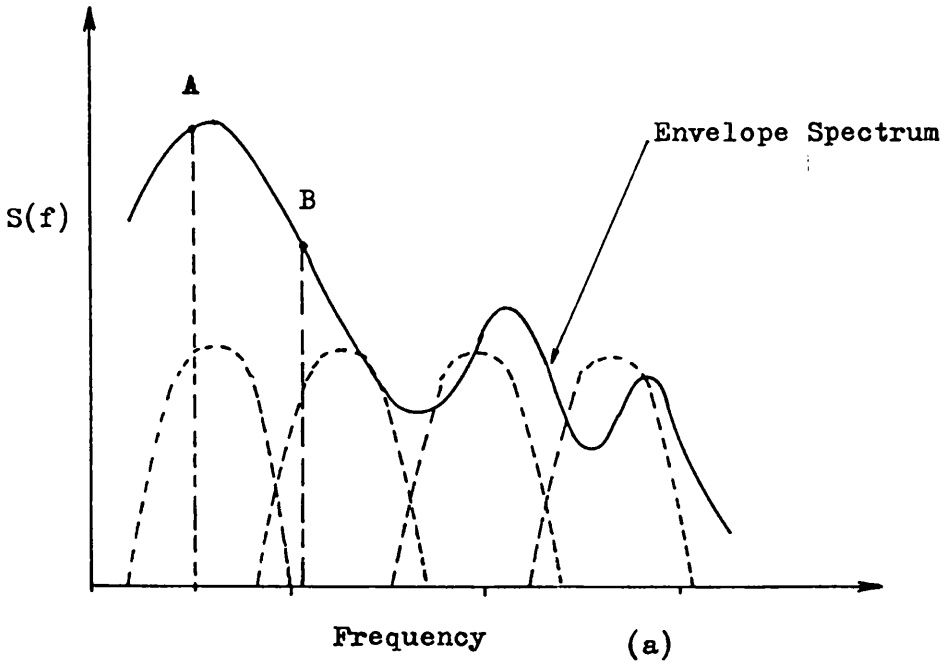
Figure 4.4 Relative Change in Voiced and Unvoiced Levels in 95% Helium , 5% Oxygen and at Different Depths.
T_v (w) : Voiced , T_u (w) : Unvoiced

4.3.5 Relationship Between the Short Time Fourier Transform of Helium Speech and Normal Speech

To process speech by an analysis synthesis technique the transfer function of the spectrum of Helium speech must be divided in the frequency domain as explained in Section 4.2. The band pass filters used for this technique can be specified initially by selecting a set which would match the transfer function of the speech production model in normal air as shown in figure 4.5a. When the transfer function is expanded for Helium Speech the parameters of those of the filters are expanded by the same factor as frequency characteristics of the vocal tract transfer function (Figure 4.5b). If the set of band pass filters used for the normal speech resolve the important parameters of this function, then the same set, with expanded characteristics, should resolve these parameters for Helium speech. Therefore each filter used to analyse Helium speech has a corresponding imaging filter used to analyse normal speech. Further their characteristics are related by the same expansion factor as that of the vocal tract transfer function.

Introducing this new assumption into the transfer function of the speech production mechanism given by equations 4.20 and 4.21 and compensating for the high frequency attenuation in the transfer function, then this equation can be rewritten as:

$$T_h(\alpha s) = \sum_{N=1}^{J} \frac{E(\alpha s)\cdot \rho_h \cdot c_h}{\rho_a \cdot c_a \cdot \alpha^2} T_a(s)\cdot H_{aN}(s) \quad \text{voiced} \qquad 4.22$$



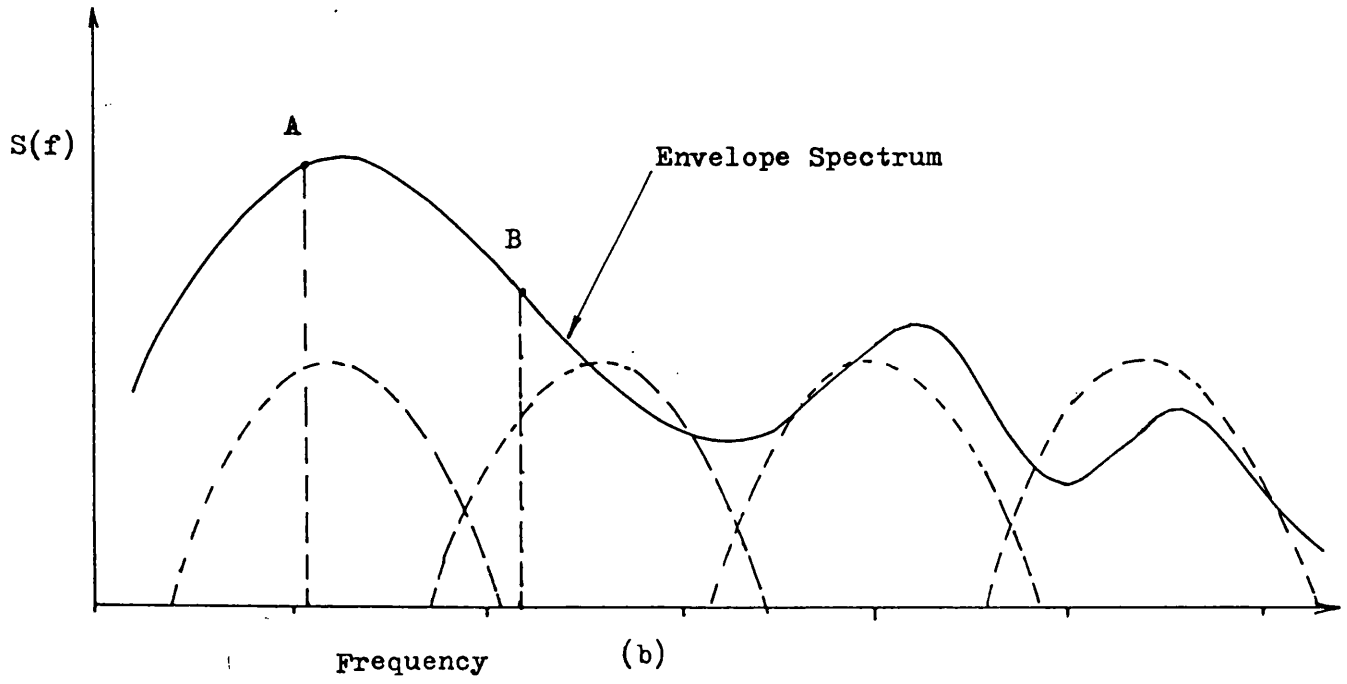


Figure 4.5 Filtering of Speech by Band Pass Filter
(a) Normal Speech (b) Helium Speech

$$T_h(\alpha s) = \sum_{N=1}^{J} \frac{E(\alpha s)\rho_h c_h}{\rho_a c_a} T_a(s)\ H_{aN}(s) \qquad \text{unvoiced} \quad 4.23$$

where $Ha_N(S)$ is the transfer function of the Nth band pass filter and $E(\alpha s)$ is an empirical expression derived to equalise amplitude distortion.

Although the expansion factor α and $E(\alpha)$ are continuous functions with frequency they can be expressed as discrete values over the frequency range. This approximation will have little effect on the processed speech due to the tolerance of the speech perception.

Denoting αs by (v) in the complex plane equations 4.22 and equation 4.23 can be rewritten as:

$$T_h(v) = \sum_{N=1}^{J} \frac{E_N \rho_h c_h}{\rho_a c_a \alpha_N^2} T_a\left(\frac{v}{\alpha}\right) \cdot H_{aN}\left(\frac{v}{\alpha}\right) \qquad \text{voiced} \quad 4.24$$

and

$$T_h(v) = \sum_{N=1}^{J} \frac{E_N \rho_h c_h}{\rho_a c_a} T_a\left(\frac{v}{\alpha}\right) \cdot H_{aN}\left(\frac{v}{\alpha}\right) \qquad \text{unvoiced} \quad 4.25$$

where E_N and α_N are the amplitude equalisation and expansion factors of the Nth band.

By taking the inverse fourier transform of equations 4.24 and 4.25 and comparing these with equation 4.8, then the steady state speech waveform at the output of speech production model will be given by:

$$Y_h(t) = \sum_{N=1}^{J} \frac{E_N \cdot \rho_h \cdot c_h}{\alpha_N \cdot \rho_a \cdot c_a} |X(\alpha_N w_N, \alpha_N t)| \cos\left[\alpha_N w_N t + \theta(\alpha_N w_N, \alpha_N t)\right]$$

voiced 4.26

and

$$Y_h(t) = \sum_{N=1}^{J} \frac{E_N \cdot \rho_h \cdot c_h}{\rho_a \cdot c_a} |X(\alpha_N w_N, \alpha_N t)| \cos\left[\alpha_N w_N t + \theta(\alpha_N w_N, \alpha_N t)\right]$$

unvoiced 4.27

Equations 4.26 and 4.27 indicate that the amplitude of the STFT of Helium speech is compressed and its magnitude is modified by a constant factor. The phase of the STFT is also compressed and its derivative now deviates around a new centre frequency.

4.3.5.1 The Relationship Between the Instantaneous Frequency of Helium and Normal Speech

The instantaneous frequency of the signal defined by equation 4.26 and 4.27, $\phi^{\cdot}(t)$ is given by:

$$\dot{\phi}(t) = w_N + \dot{\theta}(t) \qquad 4.28$$

where $\dot{\phi}(t)$ is the rate of change of the phase of the STFT with time.

To modify $\dot{\phi}(t)$ a relation is required between this frequency and the spectrum of the signal. The mean frequency spectrum W_I of the signal is given by[63]:

$$W_I = \frac{\int_0^\infty w.G(w).dw}{\int_0^\infty G(w).dw} \qquad 4.29$$

where G(w) is the power spectrum of the signal.

The power spectrum of the normal speech at the output of the analyser filter is given by:

$$G_a(w) = \sum_{m=0}^{\infty} \delta(w-mw_o).|T_a(w)|^2 |H_{aN}(w)|^2 \quad \text{voiced} \qquad 4.30$$

and

$$G_a(w) = U^2.|T_a(w)|^2.|H_{aN}(w)|^2 \quad \text{unvoiced} \qquad 4.31$$

where W_o is the pitch frequency, U^2 is the power density of the unvoiced source and δ is a delta function.

In Helium speech the sources are unchanged and equations 4.30 and 4.31 could be modified as:

$$G_h(w) = \sum_{m=0}^{\infty} \delta(w-mw_o) \cdot |T_{hN}(\alpha_N w)|^2 \qquad \text{voiced} \qquad 4.32$$

and

$$G_h(w) = U^2 \cdot |T_h(\alpha_N w)|^2 \cdot |T_{hN}(\alpha_N w)|^2 \quad \text{unvoiced} \qquad 4.33$$

The mean frequency spectrum for voiced Helium speech is obtained by substituting 4.32 into 4.29 which gives:

$$w_{Ih} = \frac{\int_0^{\infty} (\alpha_N w) \cdot \sum_{m=0}^{\infty} \delta(w-mw_o) \cdot |T_h(\alpha w)|^2 \cdot |H_{hN}(\alpha_N w)|^2 \cdot |d(\alpha_N w)}{\int_0^{\infty} \sum_{m=0}^{\infty} \delta(w-mw_o) \cdot |T_h(\alpha_N w)|^2 \cdot |T_{hN}(\alpha_N w)|^2 \cdot d(\alpha_N w)} \qquad 4.34$$

substituting the relation between $|T_h(\alpha w)|$ and $|T_a(\alpha w)|$ from

equation 4.22 into 4.34 gives:

$$W_{Ih} = \alpha_N \frac{\int_0^\infty w \sum_{m=0}^{\infty} \delta\ (w-mw_o) \cdot |\ T_a(w)\ |^2 \cdot |H_{aN}(w)|^2 \cdot dw}{\int_0^\infty \sum_{m=0}^{\infty} \delta\ (w-mw_o) \cdot T_a(w)\ | \cdot |\ H_{aN}(w)\ |^2 \cdot dw} \quad \text{voiced} \qquad 4.35$$

But the integral parts of 4.35 is the mean frequency spectrum of the output of the analysing filter in normal speech. Hence,

$$W_{Ih} = \alpha_N \cdot w_I \qquad 4.36$$

where α_N is the expansion factor at Nth filter.

Equation 4.34 indicates that the mean frequency spectrum at the output of a formant filter is expanded by a factor α_N.

Using equation 4.29, 4.31 and 4.33 the unvoiced sound instantaneous frequency is given by:

$$W_{Ih} = \alpha_N \cdot \frac{\int_0^\infty w U^2 \cdot |T_a(w)|^2 \cdot |H_{aN}(w)|^2 \cdot dw}{\int_0^\infty U^2 \cdot |T_h(w)|^2 \cdot |H_{aN}(w)|^2 \cdot dw} \qquad 4.37$$

Again the integral part represents the instantaneous frequency at the output of the Nth analysing filter. Hence,

$$W_{Ih} = \alpha_N \cdot w_I \qquad 4.38$$

It has been shown that[63] the frequency spectrum is equal to the mean instantaneous frequency $\dot{\phi}(t)$ providing the amplitude spectra have phase and amplitude symmetry, which will be satisfied if the formant frequencies are separated by band pass filter. Thus,

$$\overline{\dot{\phi}(t)} = w_I \qquad 4.39$$

To restore Helium speech formant frequencies to normality the mean frequency spectrum W_I or the average instantaneous frequency must be divided by the non linear factor α_N.

4.3.5.2 The Relationship Between the Envelope of Helium Speech and Normal Speech

The time variation of the envelope of the speech signal at the output of a wide band pass filter reflects the fine resolution of the speech spectrum (the pitch)[3, 14, 34] whilst its magnitude carries the information of the envelope spectrum of the speech. These characteristics can be demonstrated by considering the voiced model of the speech production mechanism.

For voiced sounds the steady state response S(t) of the model to periodic impulses at its input is a succession of damped sinusoids[42] given by:

$$S(t) \cong \sum_{m=-\infty}^{\infty} \sum_{N} A_N e^{-\pi B_N (t-mT_o)} \cdot \cos\left[w_N (t-mT_o) + \phi_N\right] \qquad 4.40$$

When A_N, B_N and W_N represent the formant amplitude, bandwidth and frequencies respectively.

Significant duration of these damped sinusoids is less than the pitch period[64]. Therefore the variation of the signal within a single pitch period represent the impulse response of the speech transfer function as experienced by filter N. Mathematically the impulse response $h_N(t)$ of that part of the transfer function is given by:

$$h_N(t) \simeq Y_N(t) \qquad 0 < t < T \qquad 4.41$$

In Helium speech $h_N(t)$ is more pronounced than in normal speech due to the increased formant bandwidths which reduce the interference between signals within a pitch period.

The envelope of the Helium speech signal at the output of a band filter is given by:

$$Y_{hN} = \sum_m A_N \cdot e^{-\pi\alpha B_{hN}(t-mT_o)} \qquad 4.42$$

where B_{hN} is the Helium speech formant bandwidth and A_N is the instantaneous amplitude of the Helium speech formant.

One method of modifying the amplitude of the STFT is to take the $_N$th root of the signal envelope given by 4.42. This method of modification will compress the formant amplitude and reduce the rate of decay of the envelope with time which will reduce the formant bandwidth by a factor proportional to α_N. Subjecting the envelope to $_N$th rooting will not affect the harmonic structure of the envelope signal as the signal in a pitch period will decay before the beginning of the next period. This method had been used in the past in many speech processing techniques[12] to reduce the speech

bandwidth linearly by a constant factor. This method effectively reduces each formant bandwidth. However, subjecting signals in each bandwidth to different rooting characteristics would result in non linear bandwidth compression.

4.4 CONCLUSION

The analysis-synthesis technique provides a generalised approach for speech processing. The parameters derived from the speech signal by this method reflect the physical characteristics of the speech production mechanism. These parameters are described by the short time fourier transform.

The technique can be used to compress speech signals either linearly or non linearly. One application for non linear processing is for the enhancement of Helium speech.

For Helium speech processing analysis-synthesis is the only technique with the capability of correcting the non linear frequency characteristics of Helium speech[33, 34, 38].

CHAPTER FIVE

THE REAL TIME ENHANCEMENT SYSTEM

5.1 DESIGN PHILOSOPHY

The theory of the analysis-synthesis speech processing technique has been applied to the design of a real time speech system. This system consists essentially of three major parts: the analyser, the processor and the synthesiser as shown in Figure 5.1.

The analyser resolves the speech into sixteen band limited signals. These signals are then represented over a short time interval by their envelopes and zero crossing rates. Their parameters are then dynamically modified in the processor section which consists of both envelope and frequency equalisers. The outputs of these equalisers are related to their inputs by a predetermined relation. The equaliser and analyser filters can be easily programmed to obtain optimum overall system performance. The synthesiser consists of sixteen programmable sinusoidal digital synthesisers with their frequencies and amplitudes dynamically controlled by the equalised frequency and envelope parameters. These modified signals are then summed at the output.

The system has the ability to process independently both the frequency and envelope information. The time interval over which the frequencies are processed is greater than that of the envelopes. This is in order to preserve the speech pitch which is defined by the variation of the envelopes with time and maximise the frequency resolution of the system.

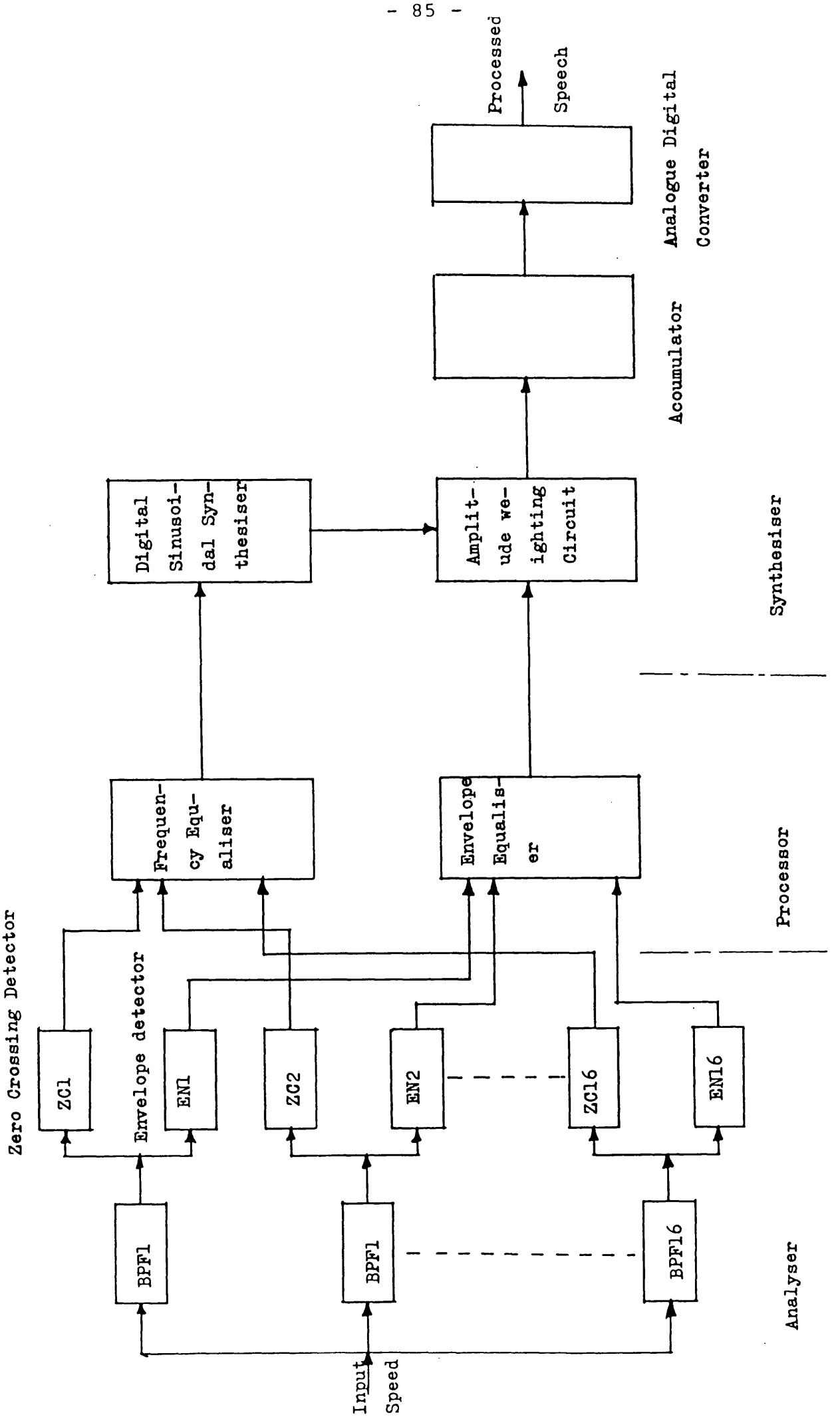


Figure 5.1 The Real Time Analysis - Synthesis Processing System.

5.2 THE ANALYSER SECTION

5.2.1 Band Pass Filters

A primary requirement is that the initial filter specification should be chosen so that their overall response should exhibit less than ± 2db over the speech bandwidth[15].

However the actual number of channel filters and thus their bandwidths will be determined by the intended speech processing technique.

These specifications are difficult to meet with conventional filters because of the inflexibility of their characteristics. However, switched capacitor filters have been chosen because of their programmability. These integrated circuit filters have the capability of producing a fourth order bandpass response, the centre frequency of which is determined by the clock frequency, whilst the bandwidths and gain are determined by two resistor values.

The circuit diagram of the fourth order bandpass filter used to analyse the speech is shown in Figure 5.2 Resistors R_1 and R_2 control the bandwidth of the first section whilst R_5 and R_7 control the gain and bandwidth of the second section.

The individual bandwidths of the cascaded sections can be calculated from the overall bandwidth of the filter. This bandwidth is given by[68]:

$$B_{in} = 0.643.B \qquad 5.1$$

where B_{in} is the bandwidth of each section and B is the overall bandwidth of the filter.

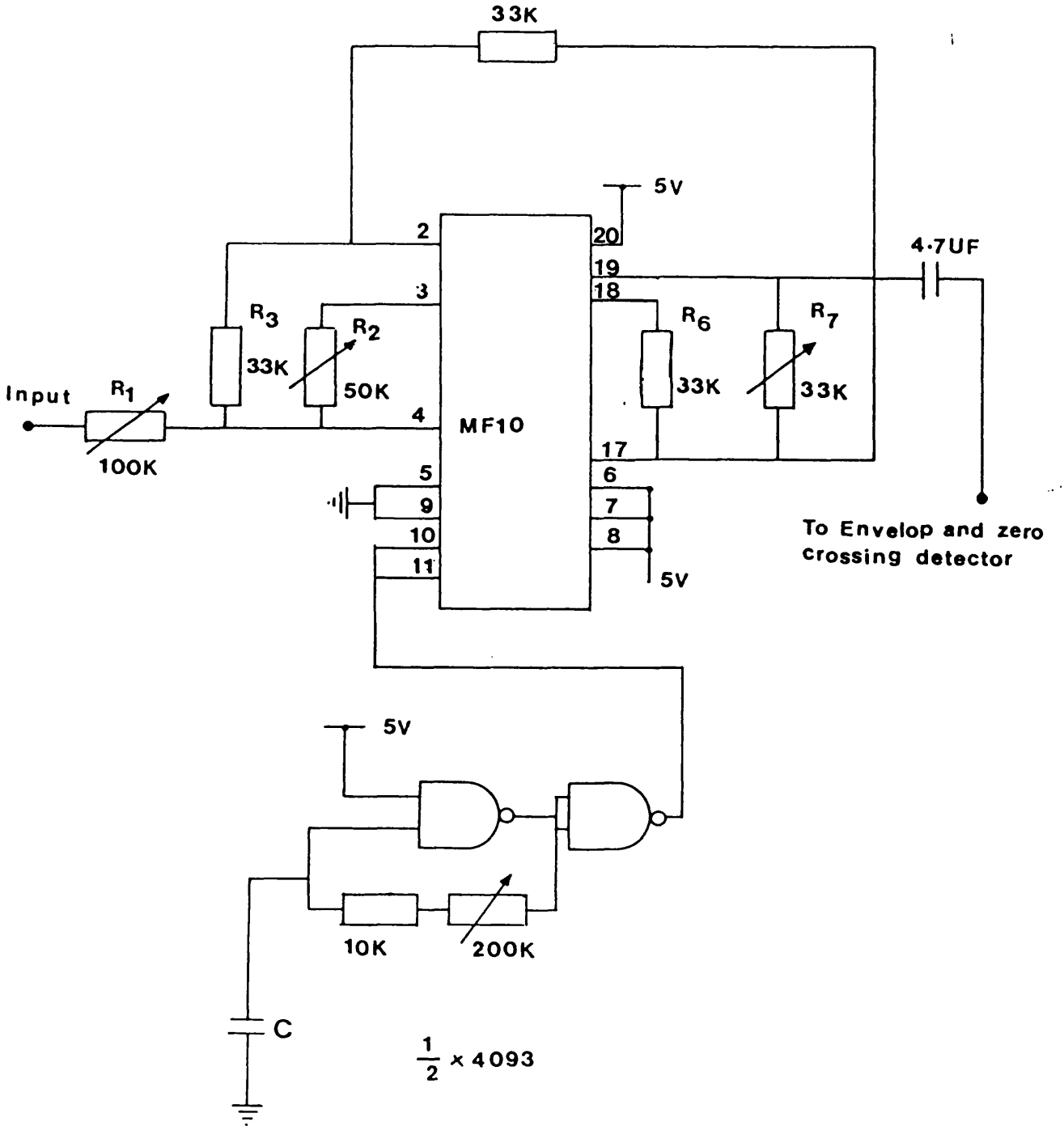


Figure 5.2 Circuit Diagram for one of the sixteen Band Pass Filters Used to Analyse the Speech.

The centre frequency of each filter is controlled by a clock frequency one hundred times greater than the centre frequency of the filter. To obtain stability cermet resistors are used to control the bandwidths and clock frequency.

It was found that adjacent filters overlappying at their -4db points and with their alternative outputs inverted gave a mimimum fluctuation in the overall amplitude response. This is shown in Figure 5.3.

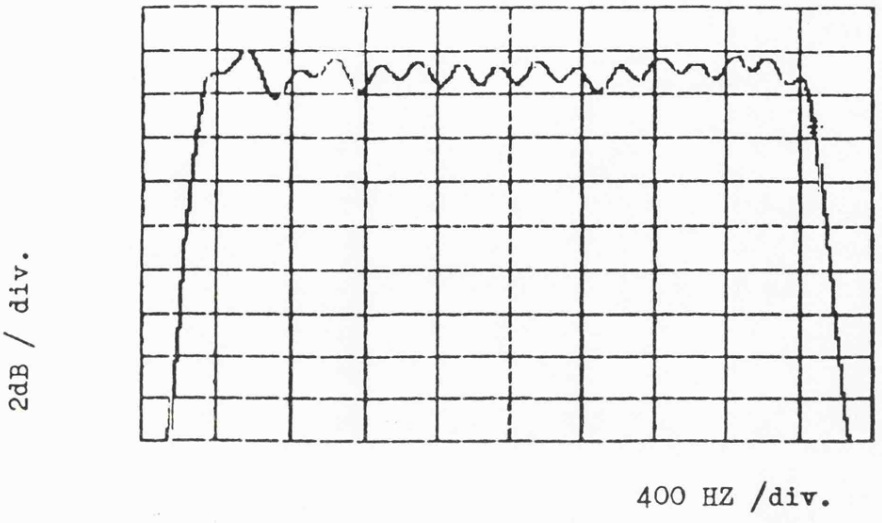
5.2.2 Envelope Detectors

To measure the amplitude of the signal in each channel sixteen envelope detectors have been constructed. The envelope detectors consist of a full wave rectifier and low pass filter (Figure 5.4), the cut-off frequency of which is determined by R_5 and C.

The outputs of the envelope detectors are available for processing through the multiplexer (I_{c1} and I_{c2}) as shown in Figure 5.5. The logic levels on ABCD select the appropriate channel, whilst the outputs from the multiplexers are converted to an eight bit digital word by the analogue to digital converter (Ic_3).

5.2.3 Zero Crossing Detectors

The zero crossings of the signal at the output of the band pass filters are counted by the circuit shown in Figure 5.6. The signal is hard limited by a high gain operational amplifier and converted to a TTL signal by the voltage comparator. The counter advances one bit each time the signal changes level. The eight bit output from



(a)

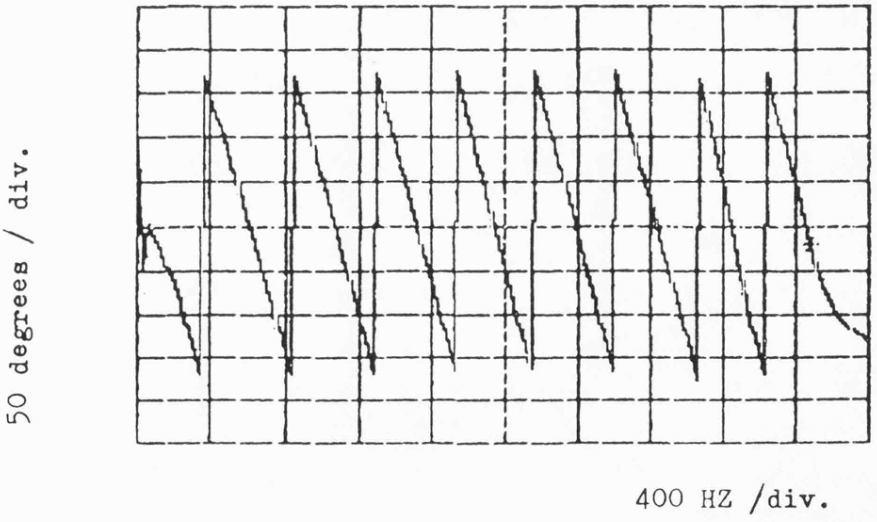


Figure 5.3 Overall Frequency Response of the Sixteen Band Pass Filters.

(a) Amplitude Response

(b) Phase Response

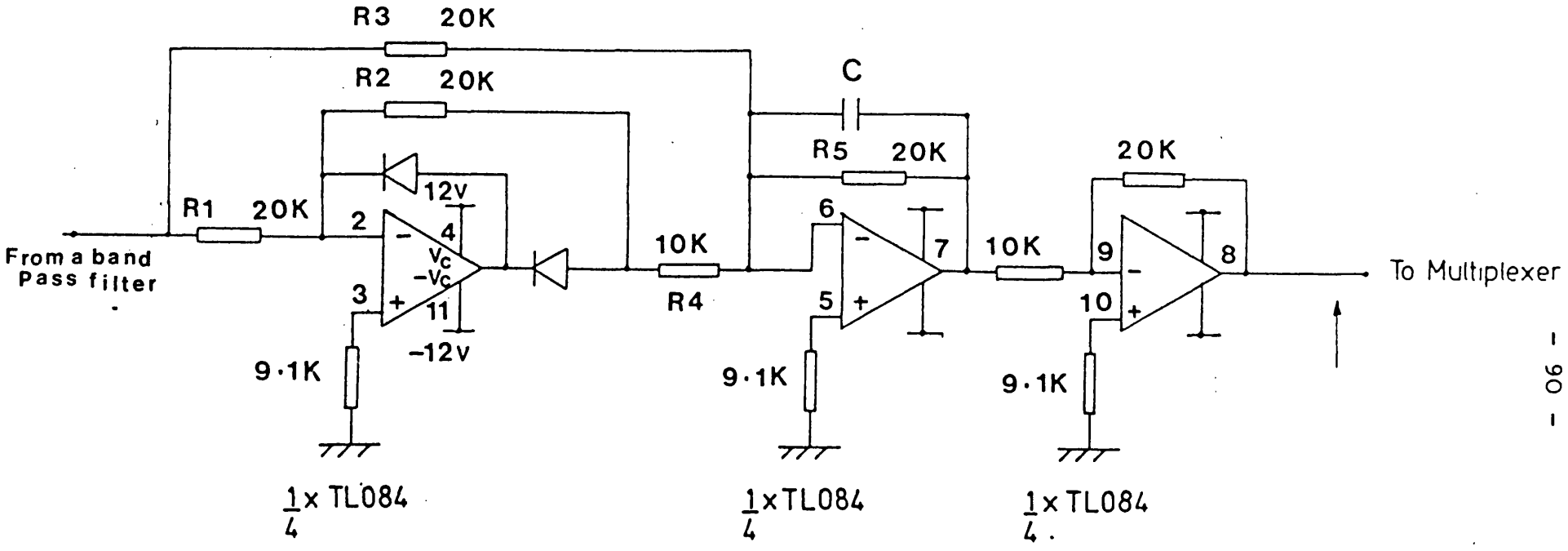


Figure 5.4 One of the Sixteen Envelope Detectors.

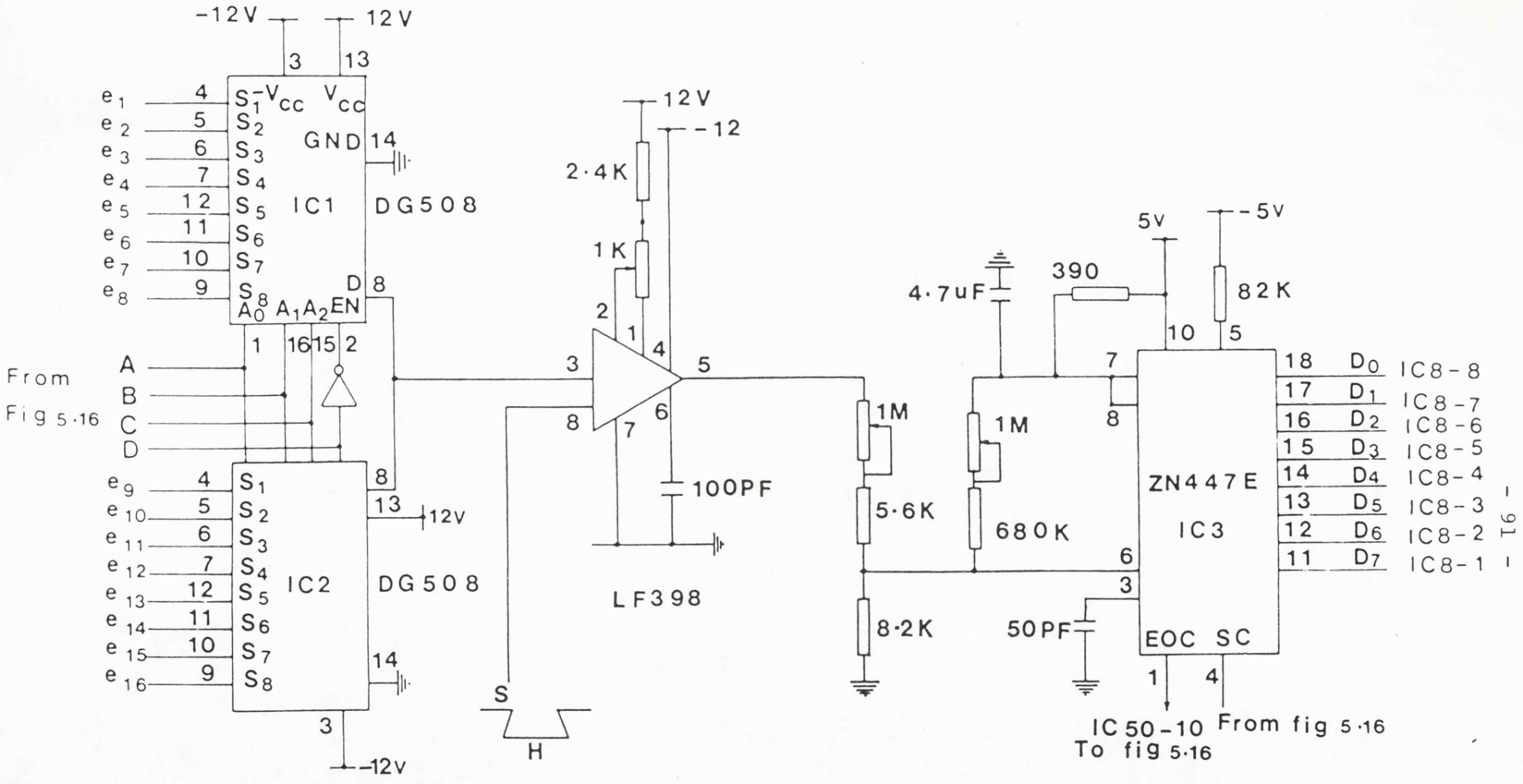

-12V
12V
IC1
DG508
GND
From Fig 5.16
A
B
C
D
IC2
DG508
12V
-12v
2·4K
1K
12V
-12
100PF
LF398
S
H
1M
5·6K
8·2K
4·7uF
390
5V
-5V
82K
1M
680K
ZN447E
IC3
50PF
EOC SC
IC50-10 From fig 5.16
To fig 5.16
IC8-8
IC8-7
IC8-6
IC8-5
IC8-4
IC8-3
IC8-2
IC8-1
Envelope Demultiplexer
Sample & Hold
Analogue To Digital Converter

Figure 5.5

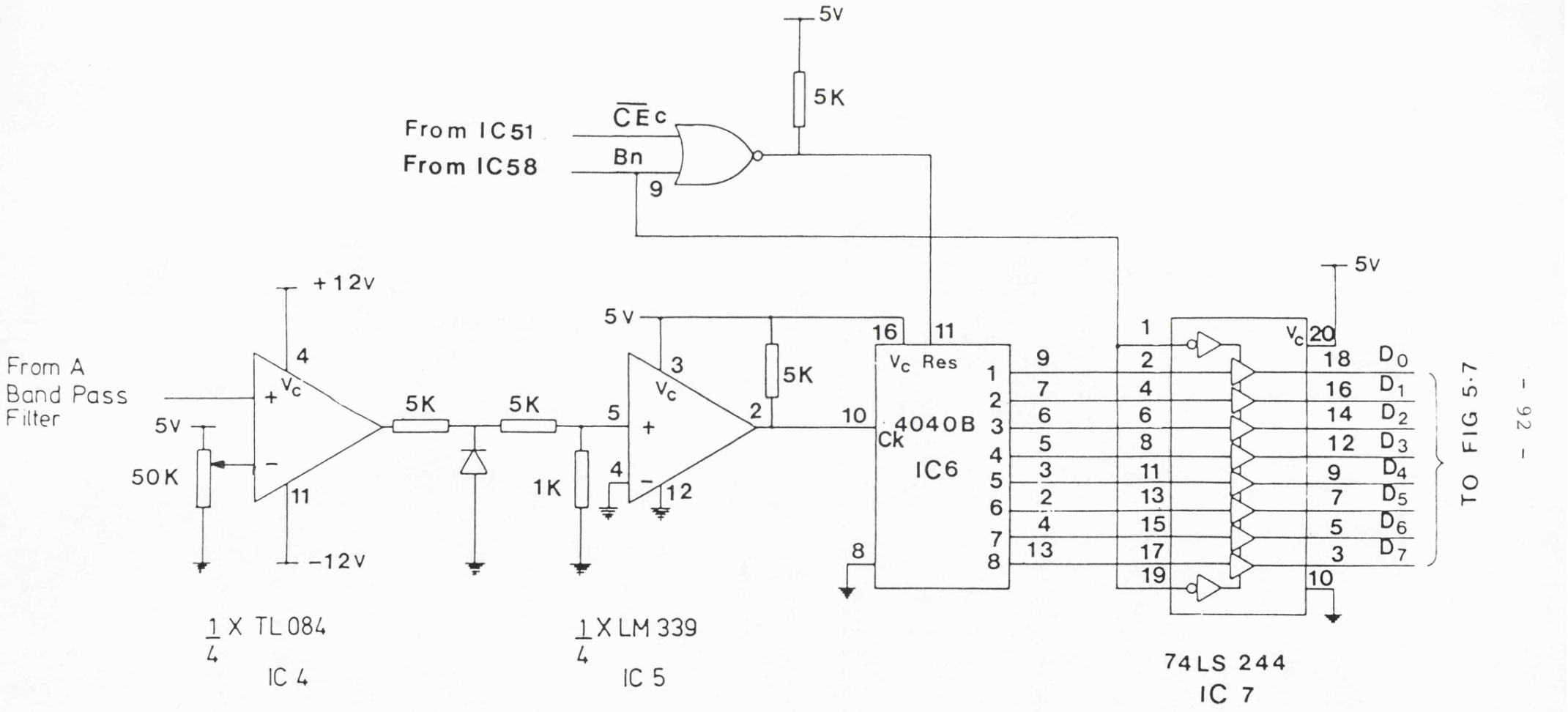


Figure 5.6 One of the Sixteen Zero Crossing Detectors.

the counter is available for processing from the TTL buffer (I_{c7}) and common latch ($I_{c5,9}$).

The resistor R_{sh} determines the signal threshold in order to minimise the effect of noise on the zero crossing measurement while the diode circuit is an over voltage protection for the comparator.

5.3 THE PROCESSOR

The processor consists of an envelope equaliser, frequency equaliser and buffer stores.

The envelopes and zero crossings are processed in different time intervals. It is therefore necessary to store the equalised envelope data to synchronise the envelope and frequency information. This maximises the resolution of the zero crossing measurements and facilitates sampling of the envelope at the maximum rate.

The envelope equaliser and buffer stores are shown in Figure 5.7. The envelope equaliser consists of sixteen kilobits of Electrically Programmable Read Only Memory (EPROM). Only eight of the eleven address lines are used as the input of the envelope equaliser, whilst the other three are used to select one of the eight available processing algorithms. The information from the envelope equaliser is buffered in one of the two sixteen kilobits memories alternately (I_{c11}, I_{c12}), whilst the other drives the synthesiser. The routing of information is determined by the state of the buffer controller Q, so that when Q is low the envelope from the analyser is stored in I_{c12} whilst I_{c11} drives the synthesiser with the modified envelope information.

The frequency processor (I_{c15} and I_{c16}) together with its buffer stores (I_{c17} and I_{c19}) is shown in Figure 5.8. The processor

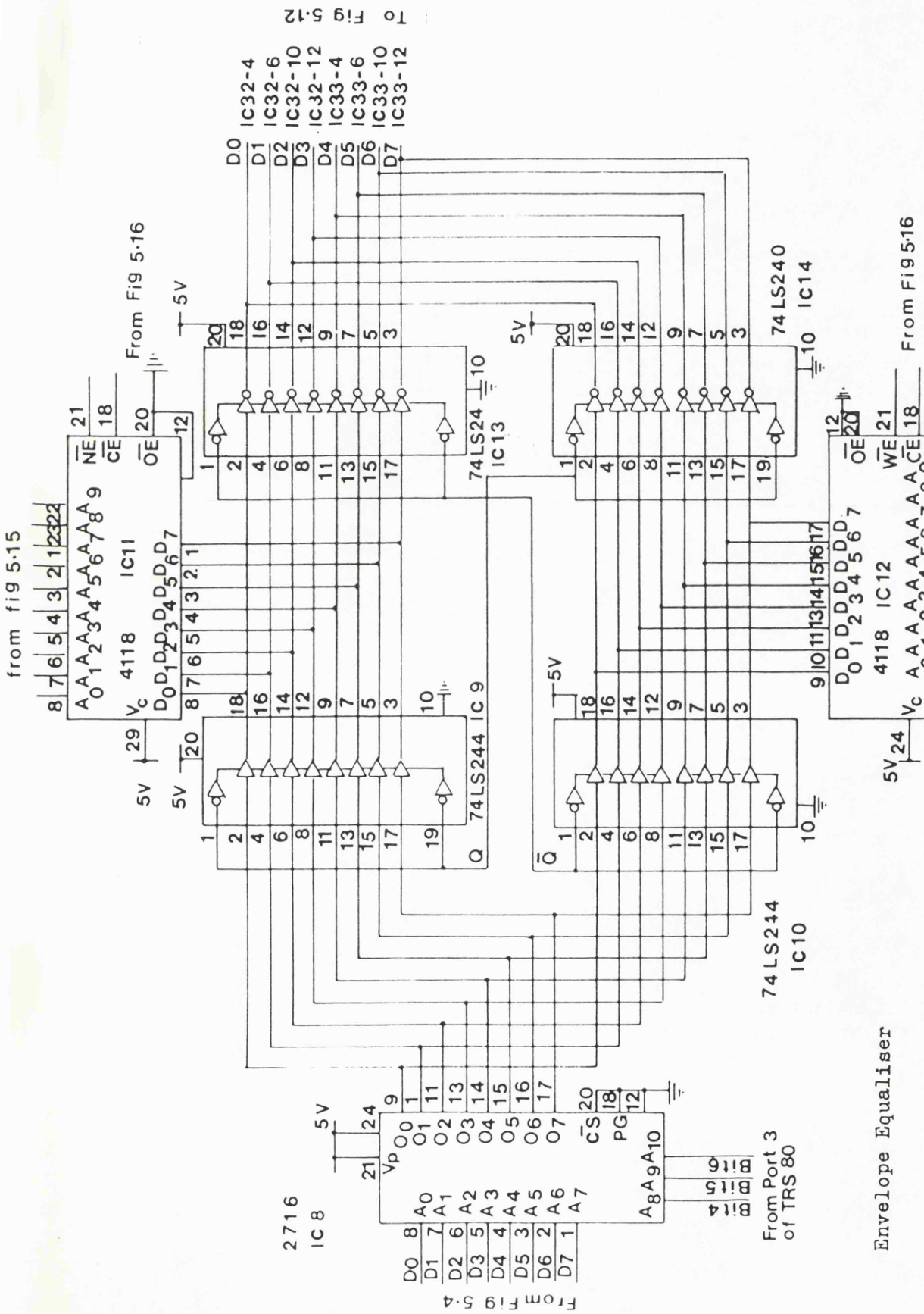


Figure 5.7 Envelope Processor

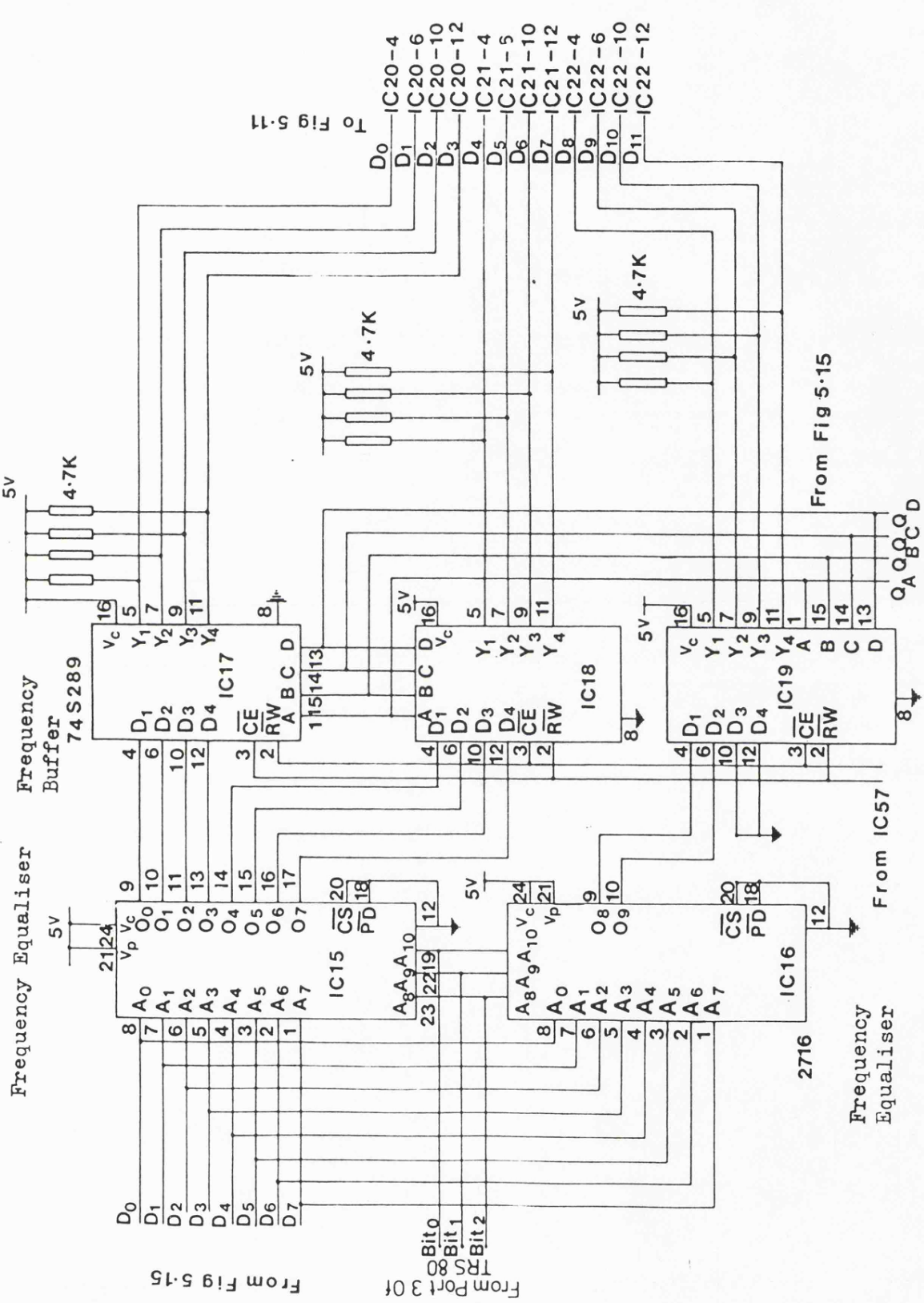


Figure 5.8 Frequency Processor.

uses a two kilobits by eight bit EPROM to provide the necessary data for frequency equalisation. The first eight bits of the address of the memory are used as an input while the data output lines represents the modified frequency. The other three address lines are controlled by external signals which allow eight different equalisation algorithms to be available from each equaliser. The output from the equaliser is a twelve bit word which is stored in sixteen by twelve bit buffers (I_{c17}-I_{c19}). Each frequency equaliser table allows two hundred and fifty six different frequencies to be modified.

5.4 THE SYNTHESISER

The synthesiser section of the system consists of four main parts; the digital sinusoidal synthesiser, the muliplier, the accumulator and the analogue to digital converter. The sinusoidal synthesiser generates simultaneously sixteen different sinusoids with their frequencies and amplitudes determined by the zero crossings and envelope information. These sinusoids are then summed digitally in an accumulator and converted to an analogue signal by the digital to analoge converter.

The maximum frequency of the synthesised sinusoids is 6.25kHz with a 6.1Hz resolution and a dynamic range of 48db with a 1db resolution.

5.4.1 Digital Sinusoidal Synthesier

The block diagram and operational principle of the digital sinusoidal synthesiser are shown in Figure 5.9 and 5.10. Samples of

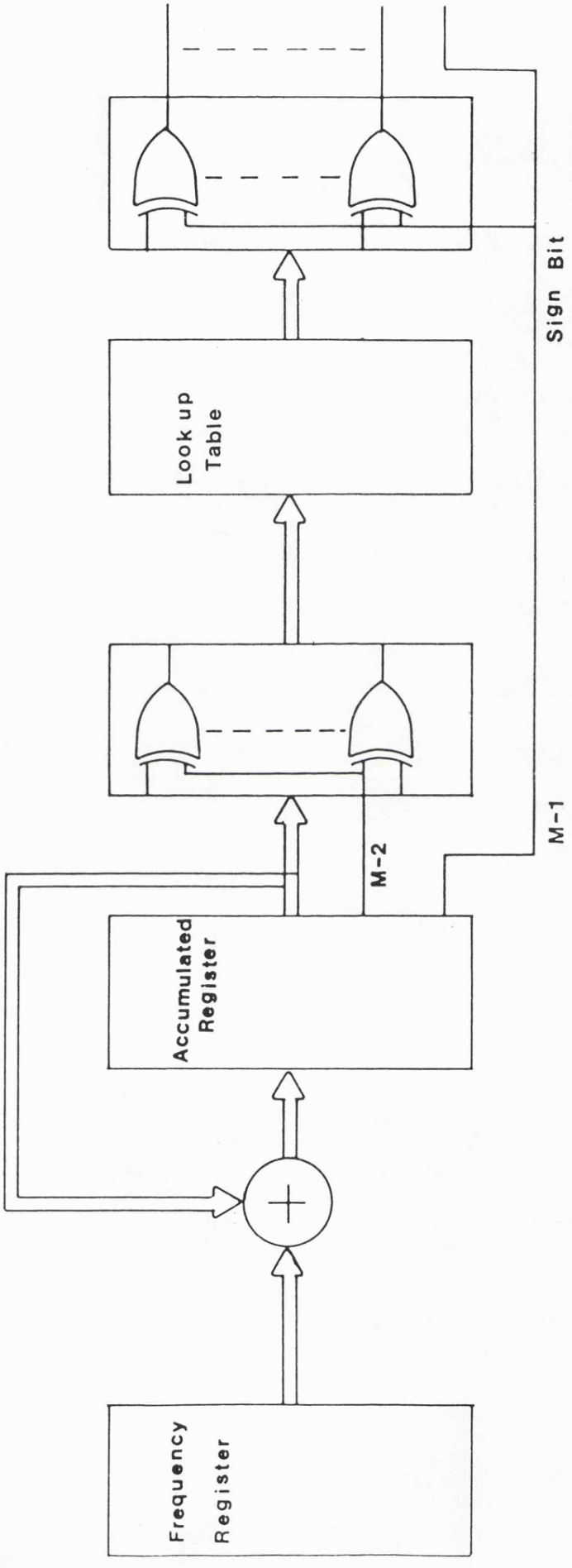


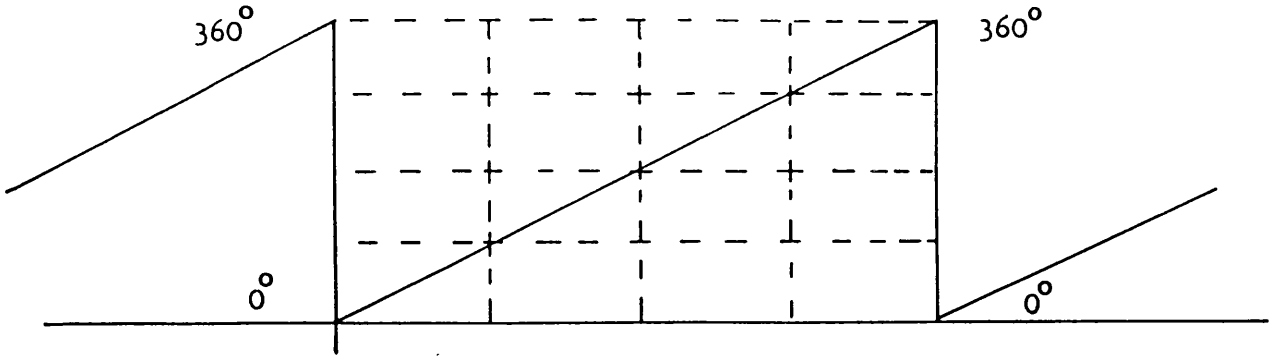
Figure 5.9 Digital Sinusoidal Synthesiser.

a quarter period of a sinusoid is stored in a look up table. The synthesiser is clocked at a constant rate while a digital number indicating the synthesised frequency is loaded into the frequency register.

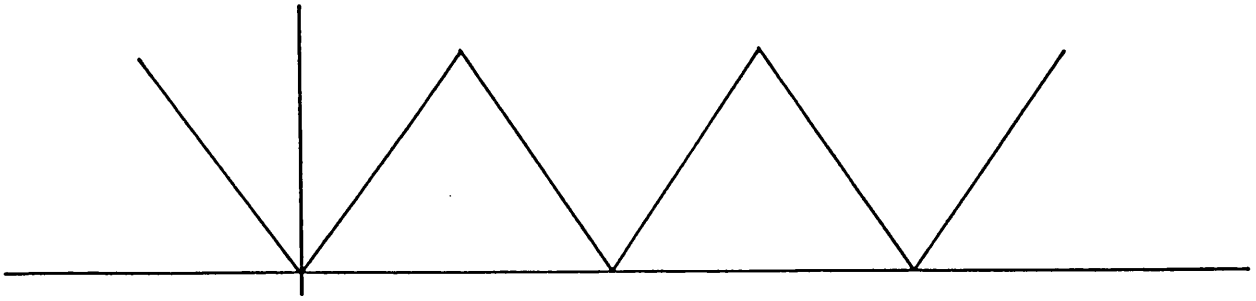
The operation of the digital frequency synthesiser can be explained by assuming that the accumulator size is (2^M-1), where M represents the total number of bits from the accumulator register. At each clock cycle the frequency register value is added to the accumulator. The output from the accumulator (the address of the sine store) selects a sample from the sine store. This sample is presented digitally at the output of the look up table.

The look up table address increases linearly with time until the entire look up table has been scanned. Then the address and the output of the look up table are at a maximum (Figure 5.10b and 5.10c). At the next clock cycle the line (M-2) from the accumulator register is high and the address of the look up table is the complement of the accumulator's register output. Hence the sine store is then scanned in the opposite direction. At the end of this scanning period a half cycle of the sinusoid will have been presented digitally to the output of the synthesiser. When the accumulator output is greater than $(2^{M-1}-1)$ the line M-2 is low and line (M-1) is high. The digital output is the complement of the look up table output, which will produce the negative half cycle of the sinusoid.

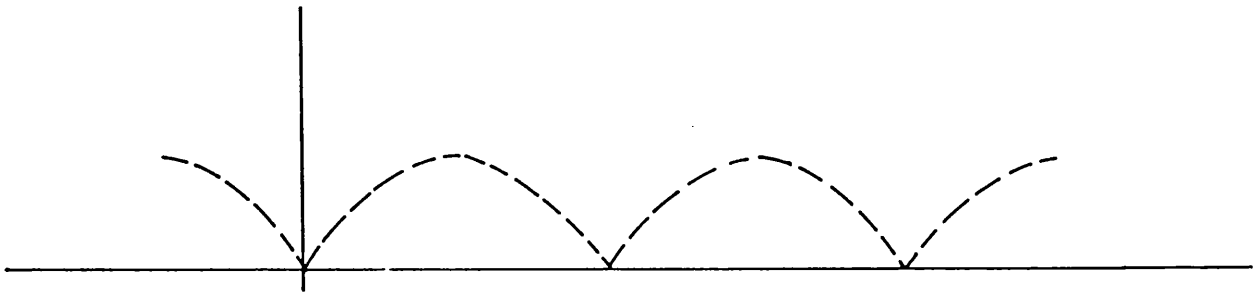
When one period is completed the accumulator is cleared and the operation repeated to produce the next period as shown in Figure 5.10c. It is clear from Figure 5.10 that different frequencies will be produced by different input values to the frequency register.



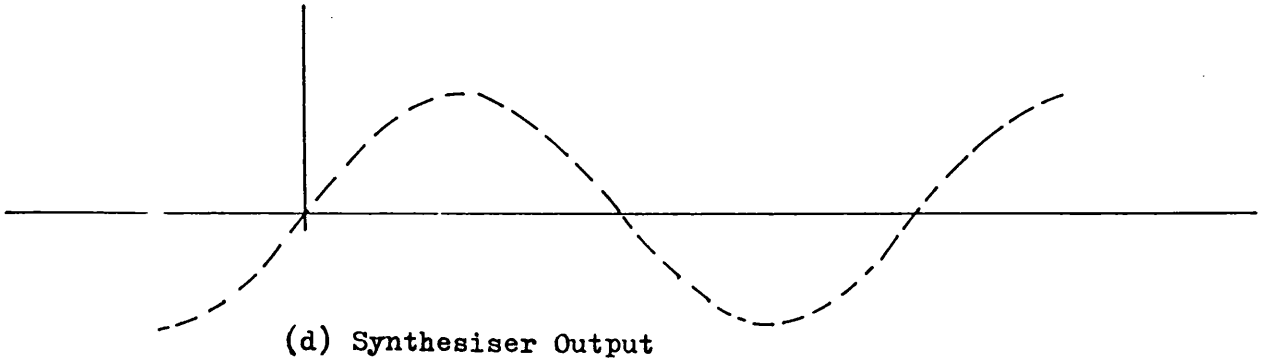
(a) Phase Accumulator



(b) Rom Address



(c) Rom Output



(d) Synthesiser Output

Figure 5.10 Filtering of Speech by Band Pass Filter.
(a) Normal Speech
(b) Helium Speech

The larger the value of this register the shorter the time needed to scan the whole look up table and the higher the frequency of the synthesised waveform. The starting phase of the sinusoidal waveform is determined by the initial value of the accumulator.

The frequency resolution of the synthesiser is determined by the accumulator size. For an M bit accumulator and a clock frequency of F_S Hz, the frequency resolution f is given by:

$$\Delta f = \frac{F_S}{2^M} \quad \text{Hz} \qquad 5.2$$

whilst the highest frequency from the synthesiser must be less than half the clock frequency according to Nyquist's theorem. The spectral purity (phase noise) is determined by the number of levels used to generate the magnitude of each sample. This can be less than the accumulator output in which cause phase jitter will be introduced into the output.

A modified sinusoidal synthesiser based on this principle has been built. The synthesiser produces simultaneously sixteen sinusoidal waveforms. The look up table address is generated by adding the value of the frequency register (Figure 5.11) to the content of the accumulator. The frequency register is a 16x12 RAM to store the frequency parameters of the sixteen sinusoidal waveforms to be synthesised. Whilst the accumulator consists of a twelve bit adder and a sixteen word store to accommodate the accumulated outputs for the sixteen channels separately.

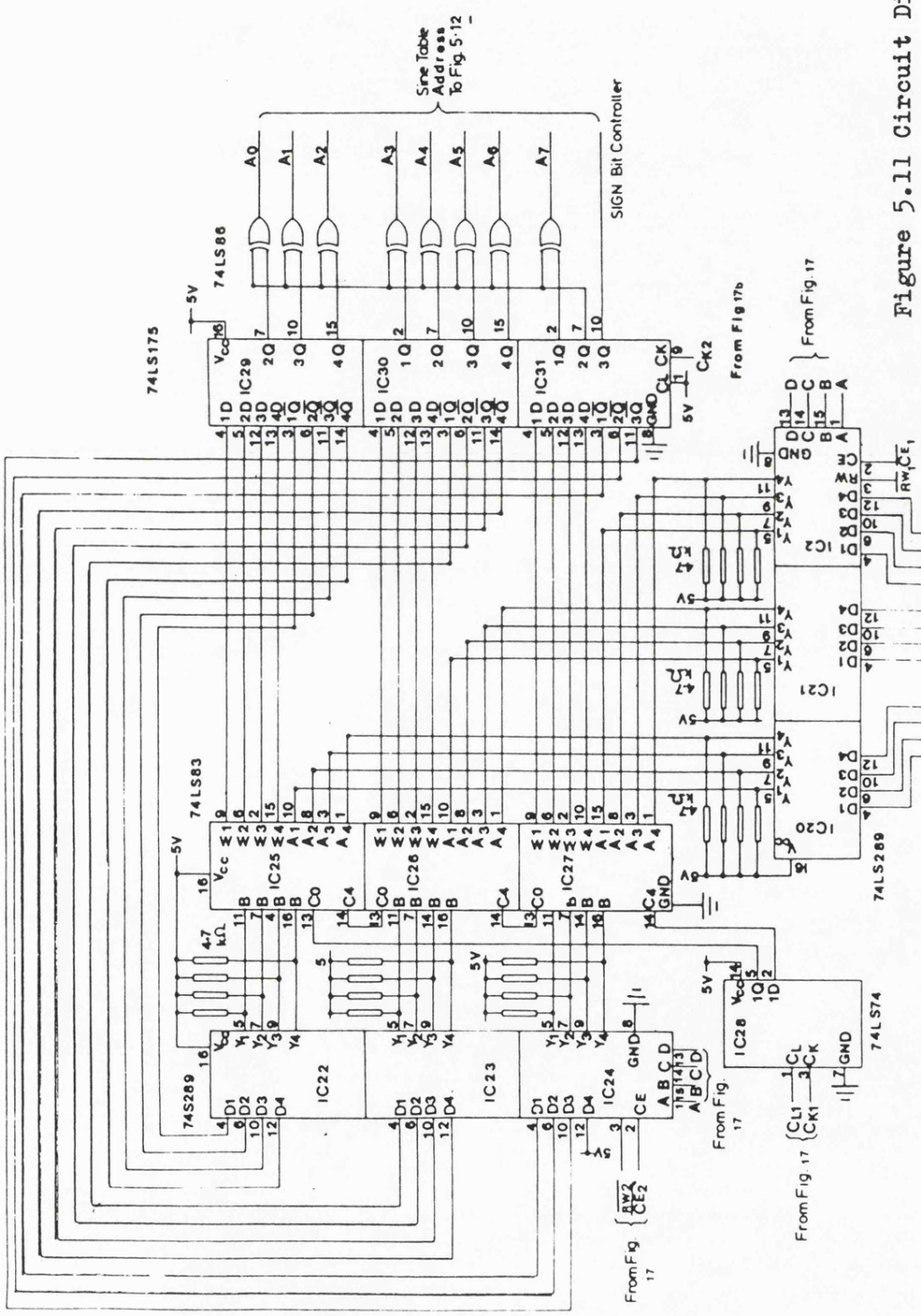


Figure 5.11 Circuit Diagram for Generation of the Look Up Table Adress.

The sinusoidal waveform will be produced as follows. The address of the frequency store and accumulator store (ABCD) is set to indicate a specified channel. The outputs from them are added together and transferred to the accumulator register. Also the output from the register is stored in the accumulator store. As before the output from the accumulator register is used to address a look up table which contains samples of a quarter period of a sinusoid. The sample from the look up table is modified by the amplitude weighting circuit (Figure 5.12). This modified output is then added to the output accumulator shown in Figure 5.13.

When a sample is added to or subtracted from the output accumulator the address of the frequency register and the accumulator store is changed to address the next channel. Then another sample is generated and stored in the output accumulator. This operation is repeated until sixteeen different samples have been generated and accumulated. Then the value in the output accumulator is converted to an analogue signal by the digital to analogue converter. The above operation is repeated at each clock cycle, after clearing the output accumulator.

The phase of each frequency generated depends on the initial value of the accumulator store (I_{C22a}-I_{C24} Figure 5.11). The phase continuity of the signal is specified by the accumulator store unless this store is reset at the beginning of the cycle.

The clock frequency of the synthesiser is 12.5 kHz which allows generation of sinusoids up to a maximum freqeuncy of 6.25kHz. With an eleven bit accumulator it is possible to generate frequencies with 6.1 Hz resolution. The quarter sinusoid is stored in 256

locations of the ROM, whilst the remaining storage of this ROM is used in the amplitude weighting circuit.

5.4.2 Amplitude Weighting Circuit

To equalise the amplitude of the sinusoids, the outputs from the look up table are multiplied by the equalised amplitude signals from the analyser section. Using a conventional digital multiplier would increase the complexity and cost of the overall system. For example with 8 bit outputs from both the look up table and envelope equaliser an 8 x 8 multiplier would be required. This multiplier would produce a 20 bit signal requiring a 20 bit output accumulator (2 x 16) to accommodate signals from all sixteen channels, also a 20 bit analogue to digital converter would also be required.

However, it has been found possible to realise the required function by the circuit shown in Figure 5.12 in which six different sets of data are stored in look up tables. Each table occupies 256 locations of the ROM. These tables contain samples from a quarter period of six sinusoidal waveforms, with 1dB difference between their maximum instantaneous amplitudes. The maximum amplitude produced from Table "0" is chosen to be 255. The maximum instantaneous amplitude of the remaining five tables are shown in Table 5.1.

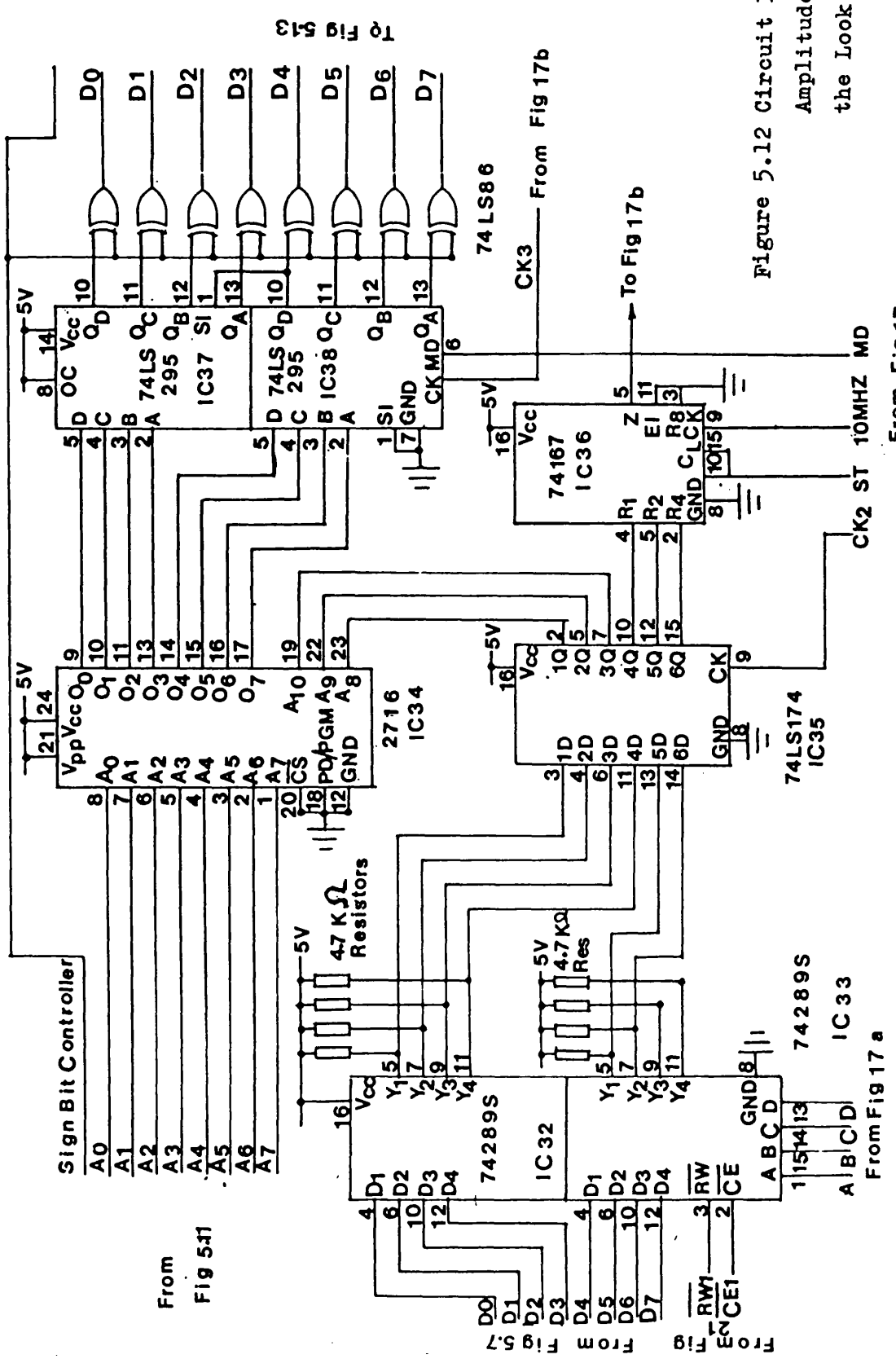


Figure 5.12 Circuit Diagram of the Amplitude Weighting and the Look Up Table.

Table 5.1. Maximum instantaneous amplitude of different tables.

Signal level in dB	0	1	2	3	4	5
Equivalent Analogue level	255	227	202	180	100	143

These six tables indicate the process that will be used to produce a dynamic range of 48dB. This is achieved by dividing the sample magnitude of different tables by multiples of two which is equivalent to reducing the signal level in muliples of 6 dB. For example dividing the samples in look up tables 0 to 5 by two is equivalent to reducing the signal level from -6dB to -11dB. Whilst dividing the samples by four is equivalent to reducing the signal level from -12dB to -17dB.

The amplitude weighting process is controlled by a six bit digital word, the first three bits represent the table number while the second three bits represent the number of the shift necessary to obtain the specified level (Table 5.1).

The weighting of the different sinusoidal signals is performed as follows. If no envelope modification is required then the data of Appendix 1 will be loaded into the envelope equaliser (Figure

5.7), the first column is the address and the last column is the equaliser data.

In operation the envelopes of each channel are modified by the equaliser before being loaded into the amplitude buffer of the sinuosoidal synthesiser (Ic 32-33). In synchronism with the frequency buffer data the amplitude of each channel is selected by address data (ABCD) and is then loaded into the amplitude latch (IC 35). The first three bits of this latch output select one of the six sine tables, the output from which is then loaded into the shift register (IC 37-38). A parallel-in shift right register is used to obtain the necessary shifts. The number of shifts of this register is controlled by the number of pulses applied to its clock terminal C_K.

The necessary pulses for the reguired shifts are provided by a decade rate multiplier (DRM). The DRM (Ic36) provides a pulse rate proportional to the digital number at its input $R_1R_2R_4R_8$. These inputs are controlled by the last three bits of the amplitude register which indicates the number of shifts required. After the required shift is completed the output from the amplitude equaliser is a sinusoidal sample modified both in frequency and amplitude.

5.4.3 Output Accumulator

At each clock cycle the modified samples from the digital frequency synthesiser are accumulated in the output accumulator of the synthesiser before being presented to the digital to analogue converter. As shown in Figure 5.13 the accumulator consists of a 12 bit digital adder/subtracter and two temporary latches. The temporary latch No. (IC44-IC45) holds the previous sample while

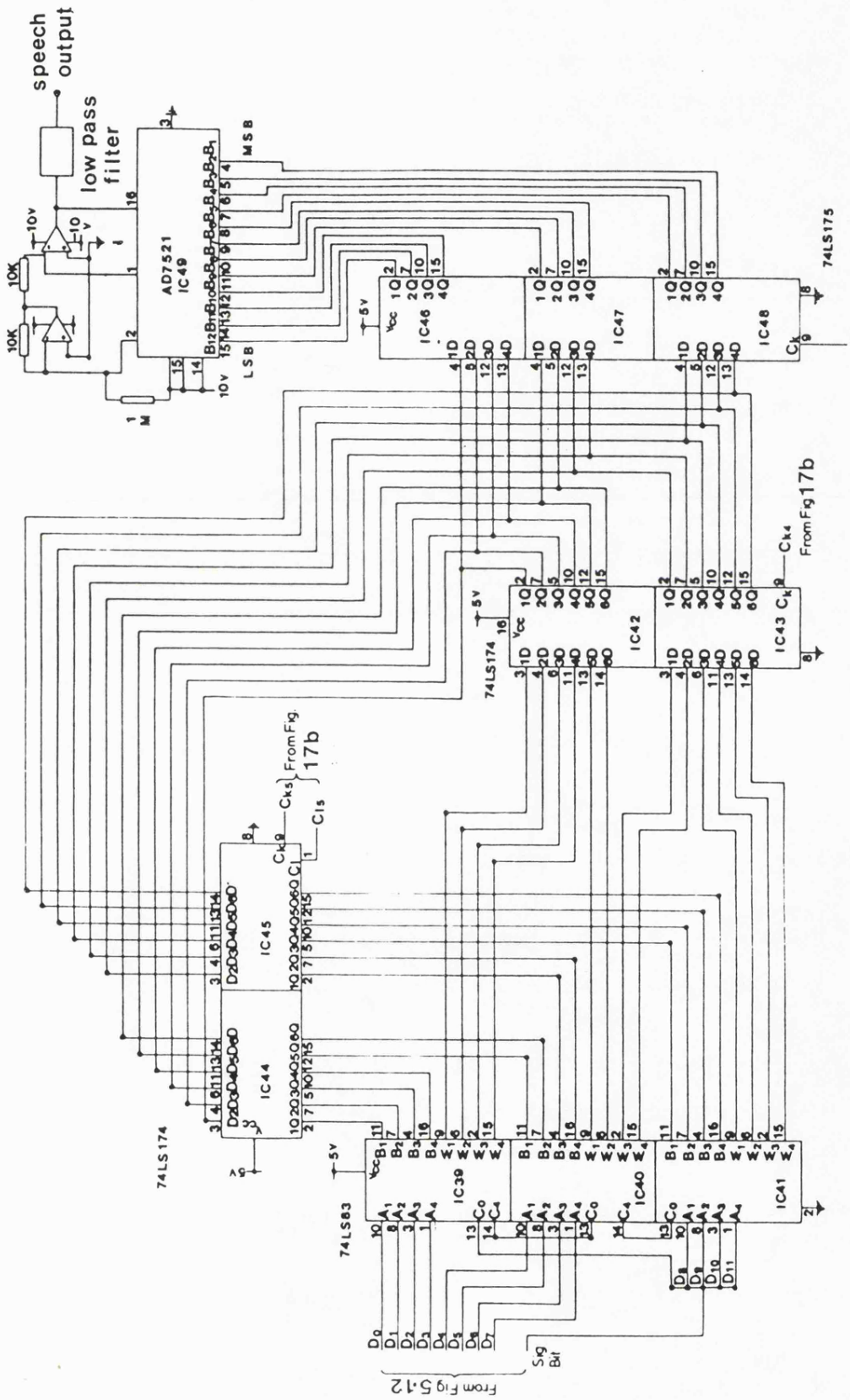


Figure 5.13 Circuit Diagram for the Output Accumulator and Analogue to Digital Convertor.

latch No. 2 (IC42-IC43) prevents any change occurring in latch No. 1 during a data transfer from the adder output to latch No. 1. The adder/subtractor mode is controlled by the add/subtract line C of I_{C39}, when C is low the input number (A's) is added to the stored number (B's), while if C is high the input number is subtracted from the stored number as shown in Figure 5.12 and Figure 5.13 a negative number is represented in 2's complement binary. The adder/substractor control signal is the sign bit from the synthesiser accumulator (I_C25-27).

The digital output from the accumulator is applied to the analogue to digital converter via a 12 bit latch. The M.S.B. of this number is inverted to match the code used by the digital to analogue converter.

5.5 TIMING SIGNALS

The timing cycle of the processing systems is shown in Figure 5.14a and Figure 5.14b. In one processing time interval the signal in each channel is represented by N envelope samples and one frequency sample, representing the rate of zero crossings. During each envelope processing interval the envelope samples are stored sequentially in one of the ampltiude buffers shown in Figure 5.7 (I_{C11}-I_{C12}). After (N-1) envelope cycles the frequencies are measured and stored sequentially in the frequency buffer shown in Figure 5.8 (I_{C17}-I_{C19}).

During the processing interval TP2 the synthesiser produces sinusoidal signals with envelopes measured at TP1 and frequency ZC_1. At the end of this interval the synthesiser reads sequentially the

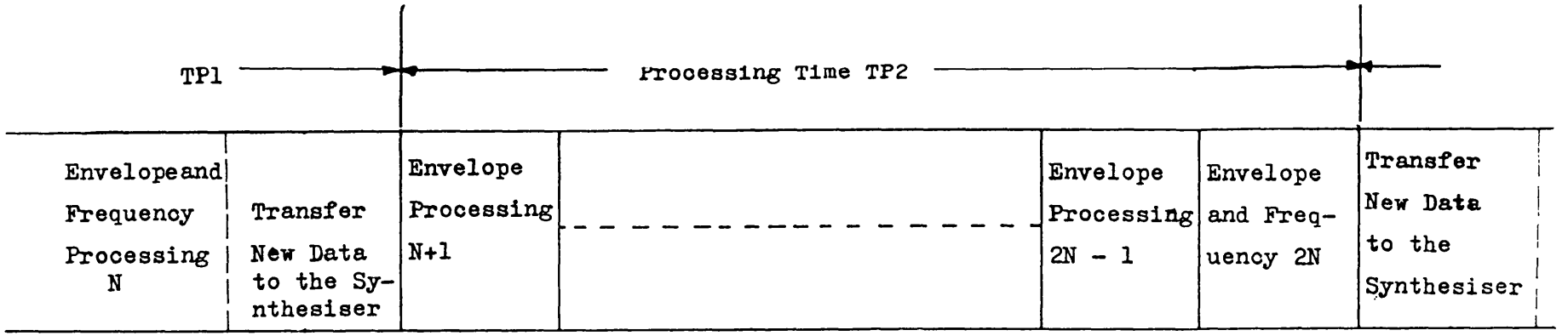
amplitude value from an amplitude buffer (I_{cl1}-I_{cl2}) while at the end of the processing interval TP_2 the synthesiser is updated with envelopes N+1 to 2N and frequency ZC2. During the transfer of the new data the synthesiser pauses before the next cycle. Since the average rate of zero crossings depends on the processing time interval different frequency modifying tables will be required.

Three types of control signals provide the system with the facility shown in Figure 5.14. They are the analyser control signals, the update control signals and the synthesiser control signals. The circuits to provide this facility is shown in Figure 5.15, Figure 5.16 and Figure 5.17.

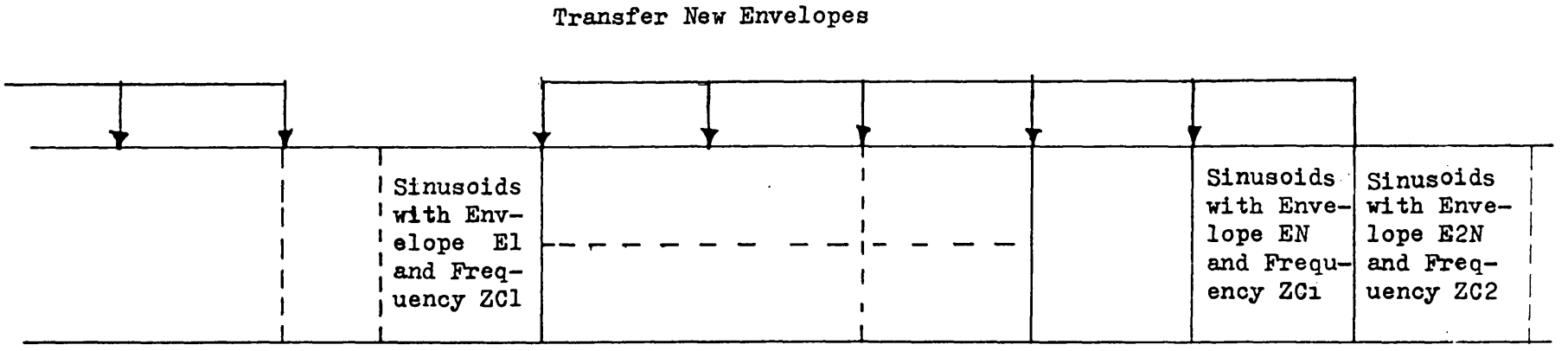
5.5.1 Analyser Control Signals

The analyser control signals are generated by a computer and from purpose built hardware. These signals produce the necessary information for selecting channels, and transferring the amplitude and frequency parameters to their appropriate buffers. They also produce commands to convert the analogue amplitudes into digital form in the analogue to digital converter. However at the end of each processing interval they provide the necessary signals to transfer this frequency and amplitude information into the synthesiser. The necessary signals for controlling the duration of the main processing interval (TP1) and the envelope processing intervals (TP2) are determined entirely by software.

The timing diagram for the (N-1)th envelope processing interval (TP2)is shwon in Figure 5.18 whilst the timing diagram for the 2Nth processing interval is shown in Figure 5.19. The limiting



(a)



(b)

Figure 5.14 General Timing Cycles of the System.

(a) Analyser Timing Cycle (b) Synthesiser Timing Cycle

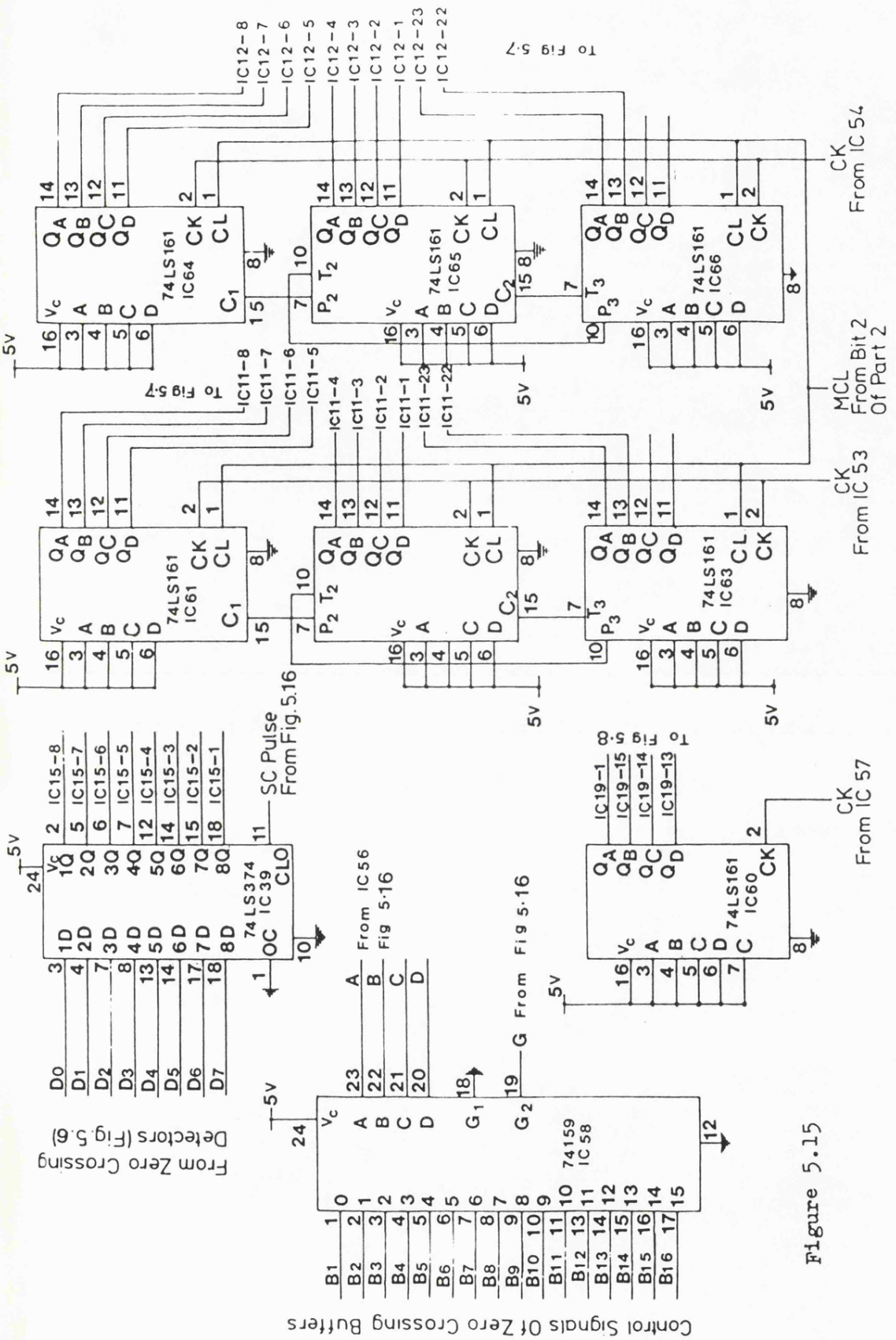

Control Signals Of Zero Crossing Buffers
From Zero Crossing Detectors (Fig 5.6)
74159 IC58
74LS374 IC39
74LS161 IC64
74LS161 IC65
74LS161 IC66
74LS161 IC61
74LS161 IC63
74LS161 IC60
From IC56 Fig 5.16
G From Fig 5.16
SC Pulse From Fig. 5.16
To Fig 5.7
To Fig 5.8
CK From IC 54
CK From IC 53
CK From IC 57
MCL From Bit 2 Of Part 2
5V

Figure 5.15

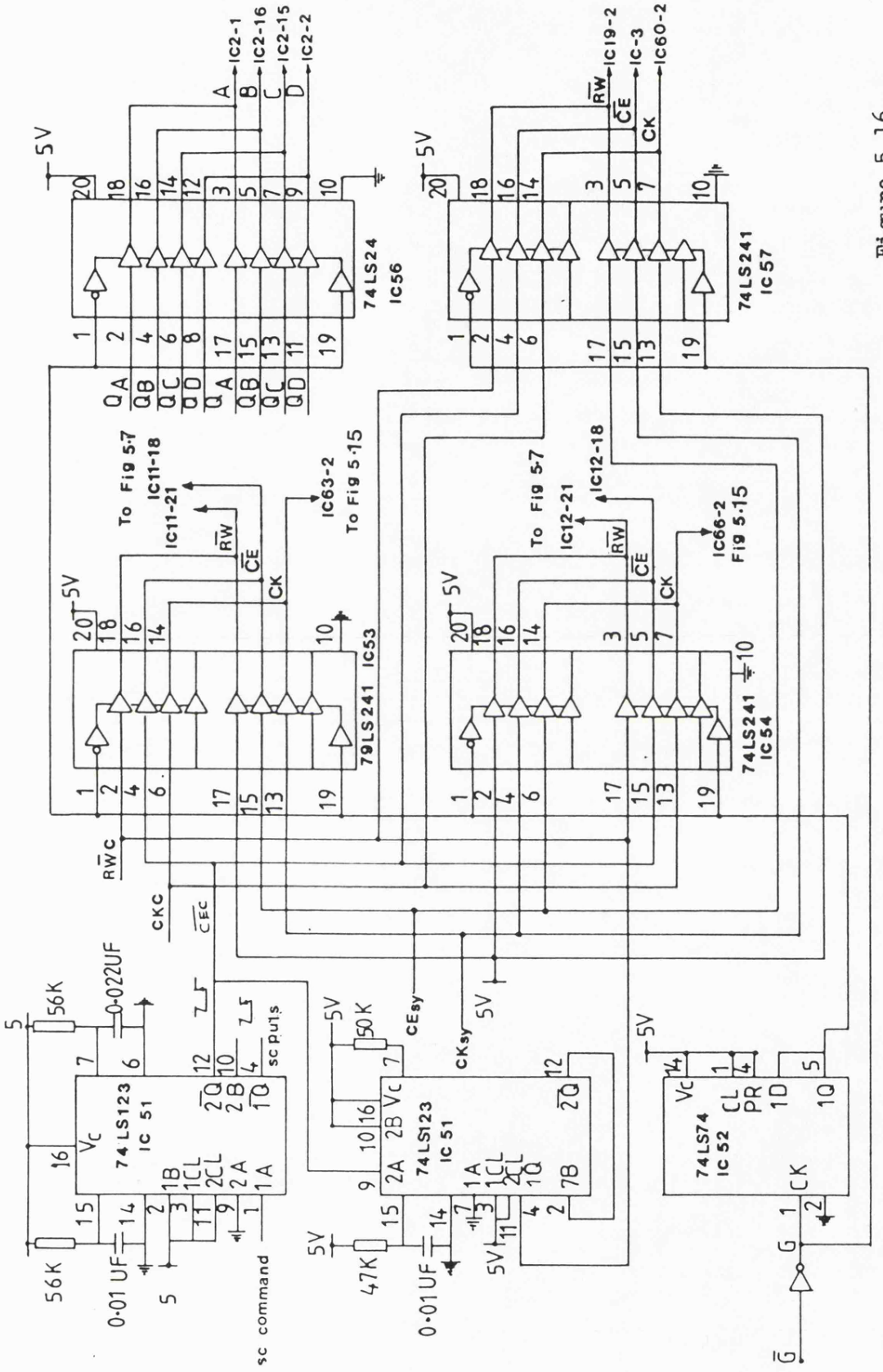

74LS24
IC56
74LS241
IC57
79LS241 IC53
74LS241
IC54
74LS123
IC 51
74LS74
IC 52
To Fig 5-7
To Fig 5-15
IC11-18
IC11-21
IC12-18
IC12-21
IC63-2
IC66-2
IC2-1
IC2-16
IC2-15
IC2-2
IC19-2
IC-3
IC60-2
56K
0-01 UF
0-022UF
47K
50K
sc command
sc puls
CKC
CEC
RWC
CEsy
CKsy
5V

Figure 5.16

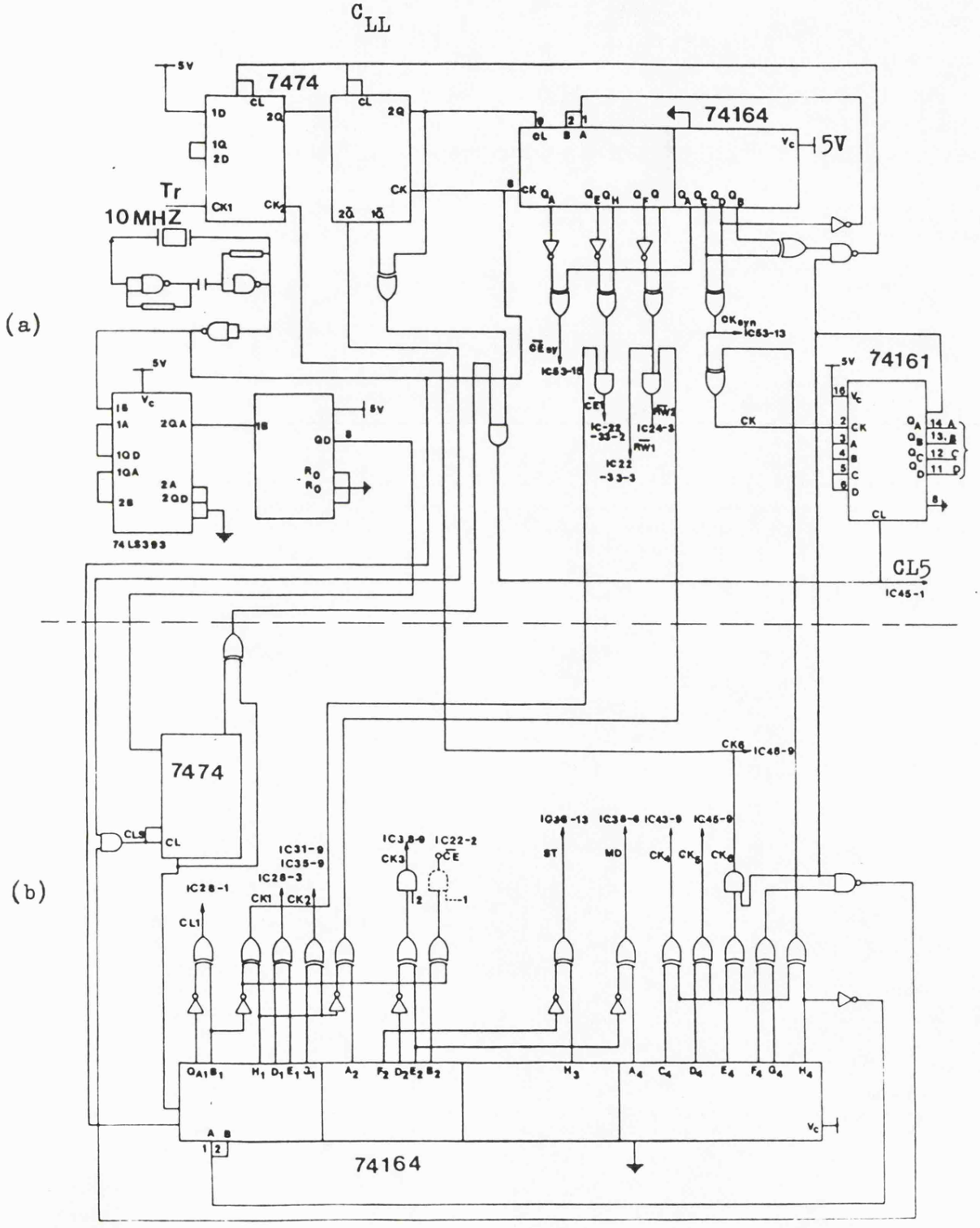


Figure 5.17

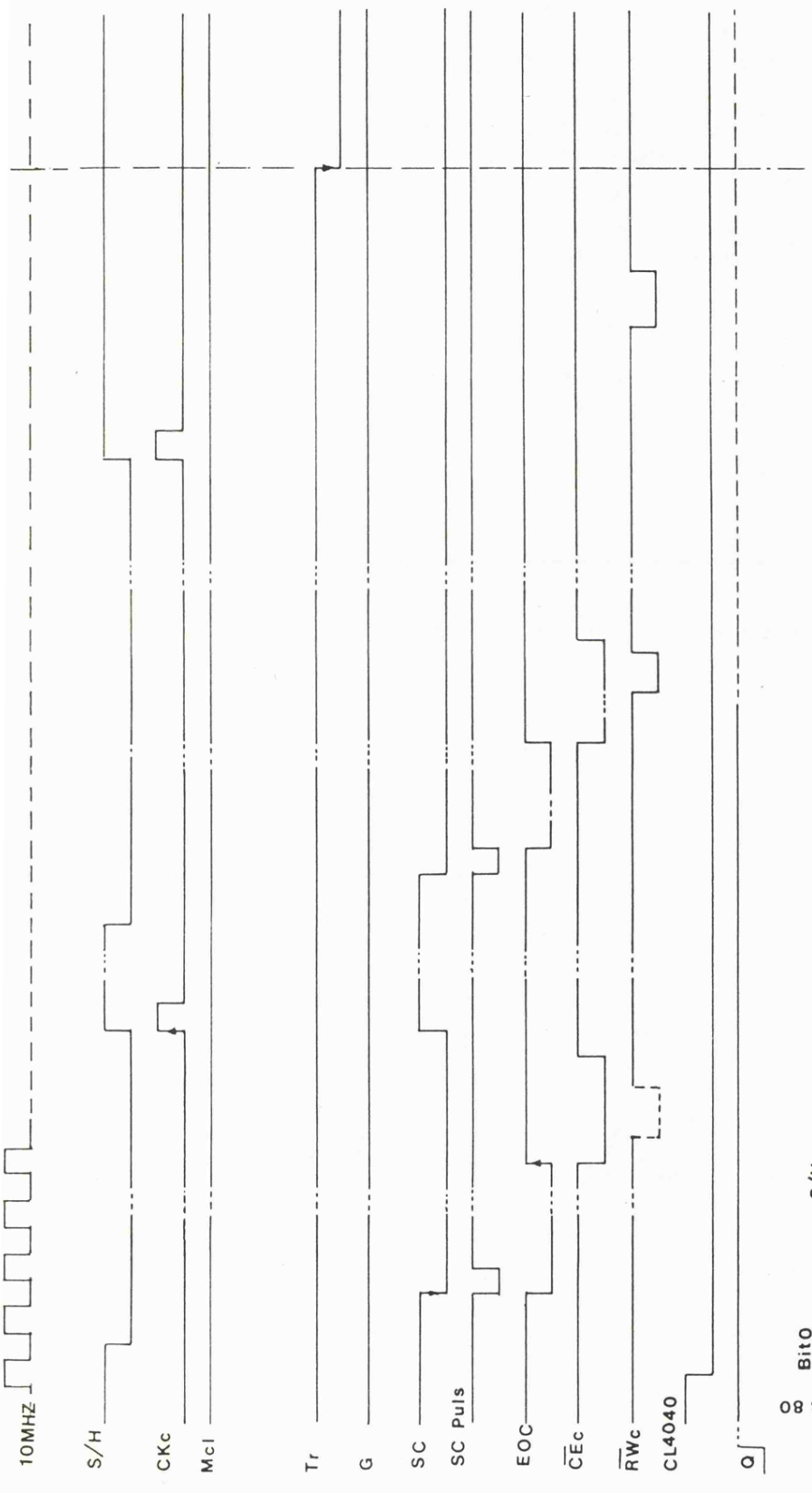


Figure 5.18 Timing Diagram for One of the (N-1)th Processing Interval.

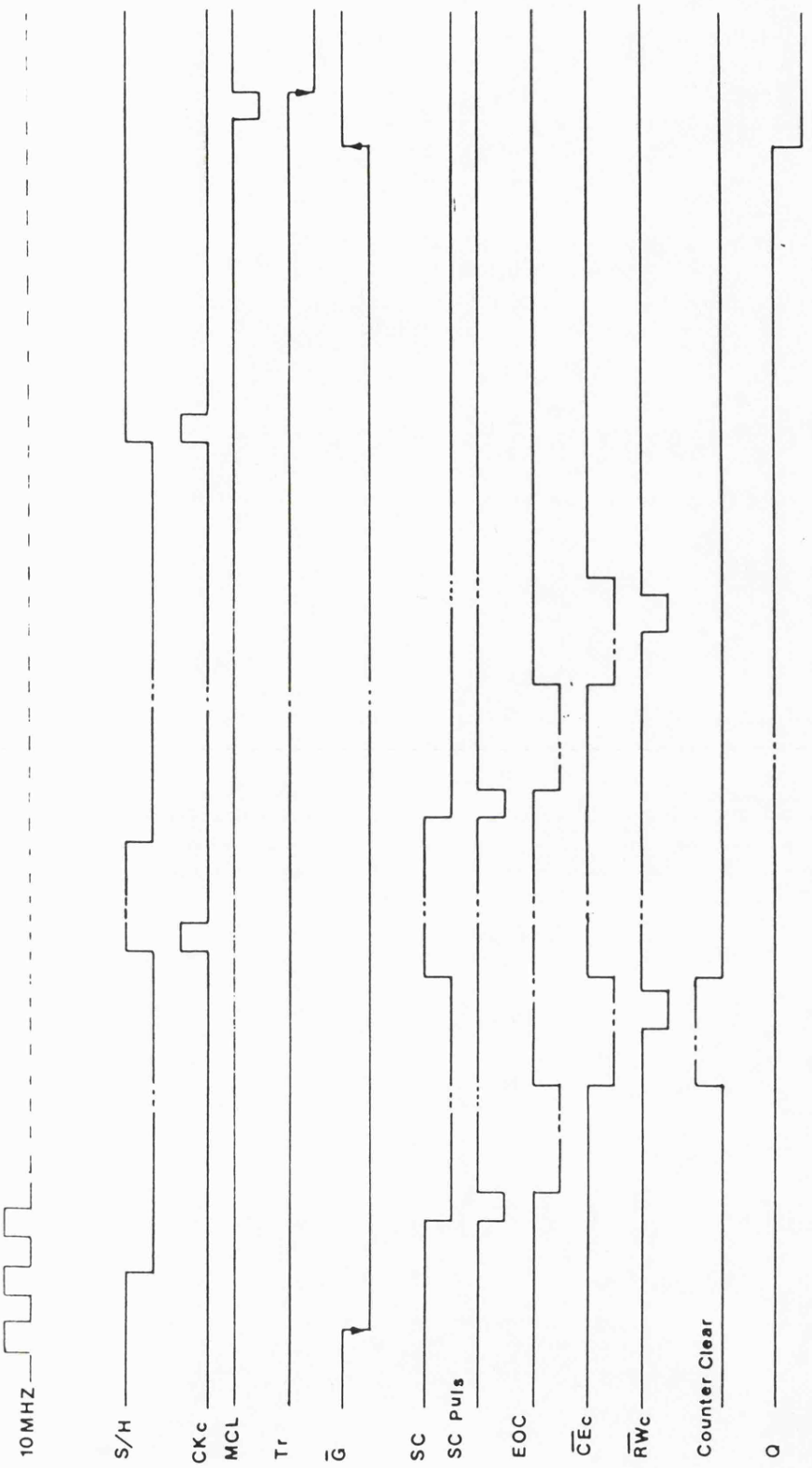


Figure 5.19 Timing Diagram for the Processing Interval (2N)

lengths of these cycles depend on the speed of the computer and of the analogue to digital converter.

5.5.2 Up Date Control Signal

Each time the transfer signal changes state from high to low the up date circuit, shown in Figures 5.17a, produces a set of signals which prevent the synthesiser from generating the new sinusoid cycle. Also it allows sixteen new amplitudes and frequencies to be loaded into the digital synthesiser.

To minimise the interruption in the output signal, due to this pause, the update cycle starts after the samples for all the channels have been generated. The timing diagram for the up date cycle is shown in Figure 5.20. However, this cycle takes 32.7 microsecond to load sixteen new frequencies and amplitudes into the synthesiser.

5.5.3 Synthesiser Control Signals

The synthesiser control signals control the digital frequency synthesiser, the amplitude weighting circuit, the output accumulator and the digital to analogue converter.

The circuit which produces these signals is controlled by a 12.5 kHz clock which is derived from a 10 MHz crystal oscillator as shown in Figure 5.17b. Each time this clock changes state the synthesiser control signals produce the necessary input to the synthesiser to generate sixteen digital sinusoidal samples and present them to the digital to analogue converter.

The timing diagrams of these signals are shown in Figure 5.21 whilst the circuit which generates them is shown in Figure 5.17b.

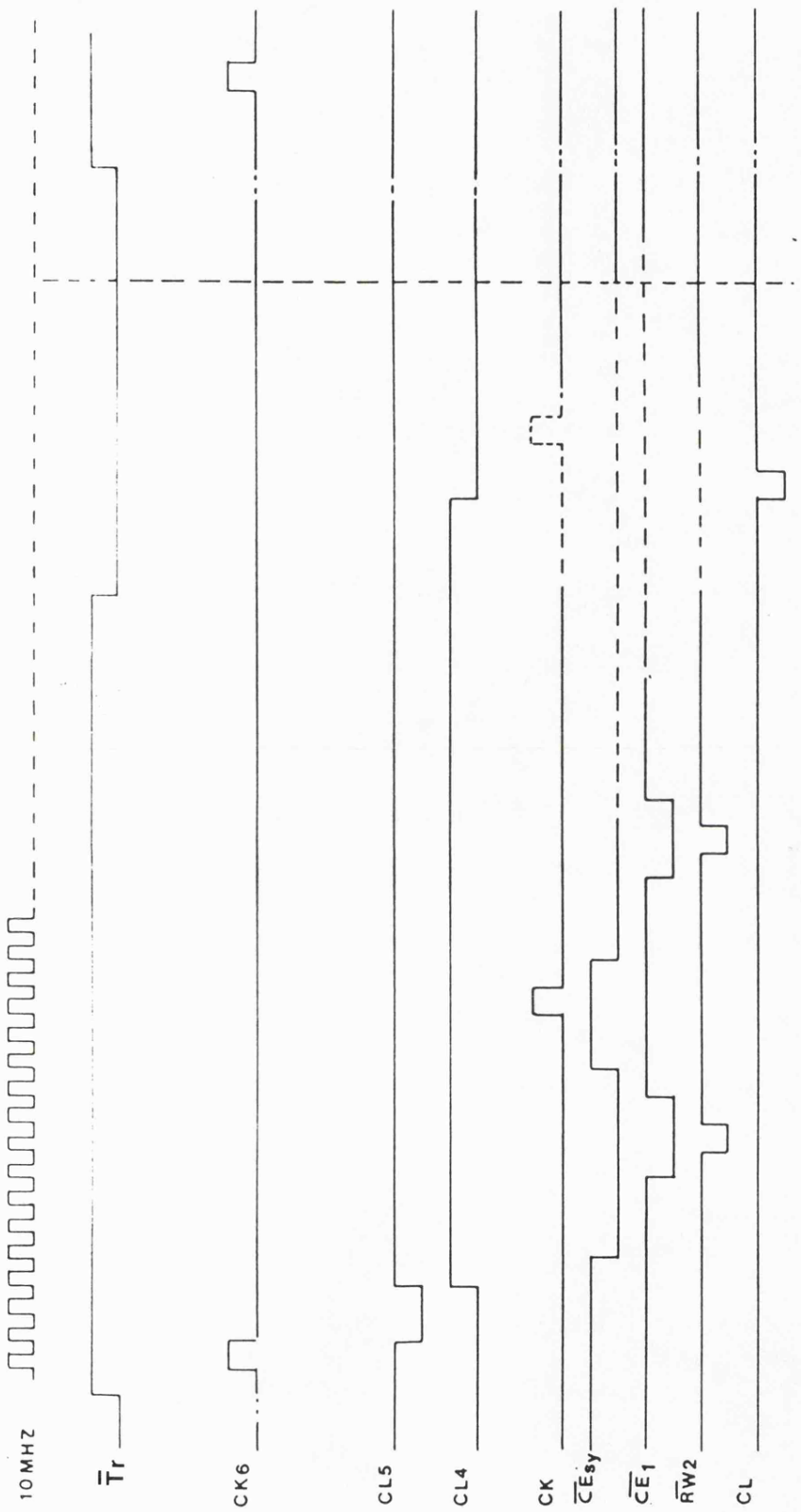


Figure 5·20 Timing Diagram for the Update Cycle.

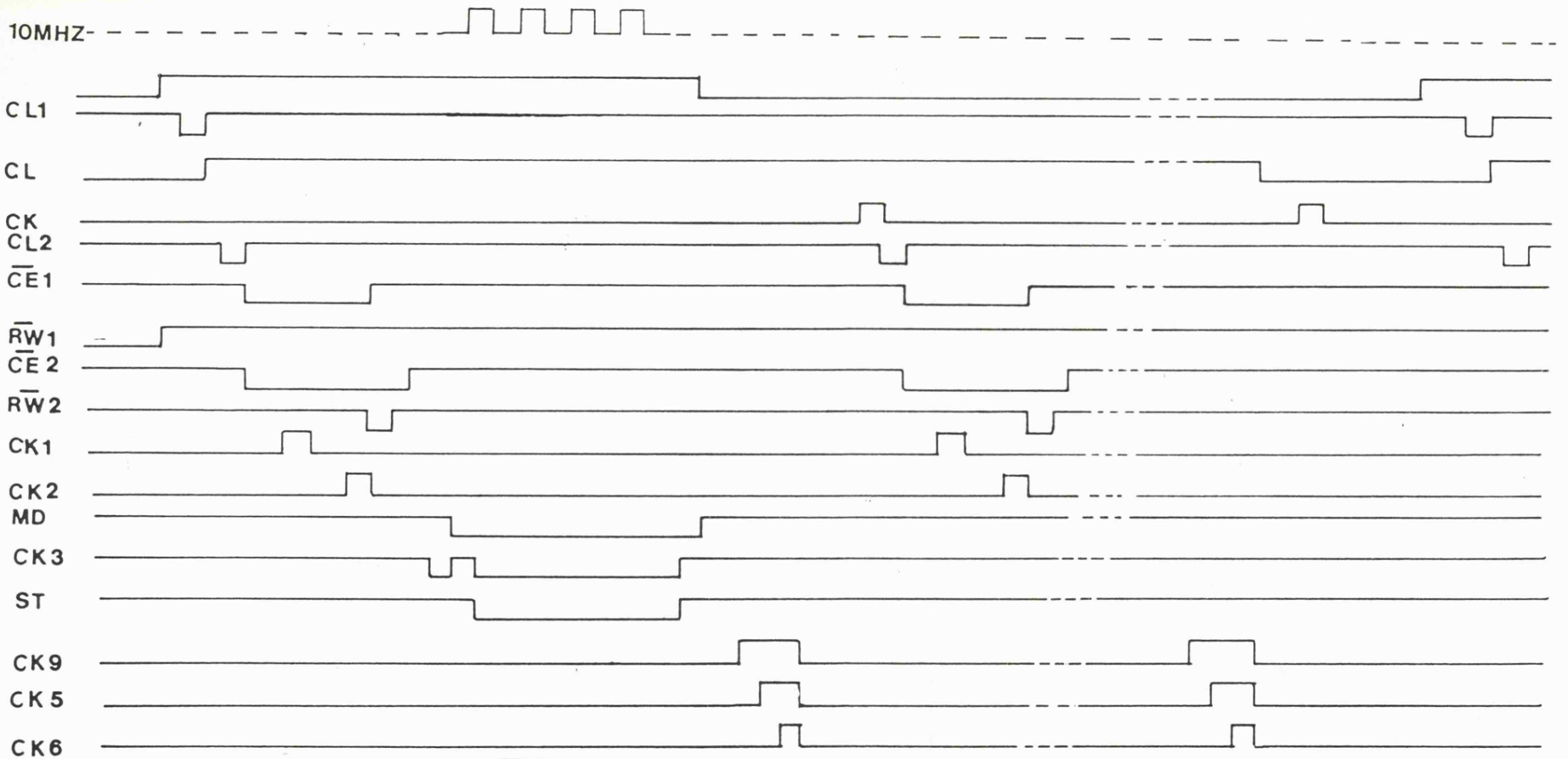


Figure 5.21 Timing Diagram for Synthesiser Cycle.

the time required to generate a sample is 3.2μsec or 51.5μsec to generate sixteen samples. However the synthesis cycle is 80μsec. The sample amplitude multiplication takes 1 μ sec because the binary rate multiplier produces inregular pulses. This generation of irregular pulses takes a maximum of 800 μsec to produce the eight required pulses.

CHAPTER SIX

OVERALL SYSTEM EVALUATION

6.1 INTRODUCTION

Generally the assessment of the ability of a speech processing system to produce an intelligible output requires an evaluation algorithm[3, 66]. This is usually achieved by counting the number of discerete speech units correctly recognised by a listener. The procedure by which the intelligibility is measured is known as an articulation test[67]. Typically, a speaker will read lists of either syllables, words or sentences to a group of listeners and the percentage of items correctly recorded by the listener will be the articulation score. This percentage is then taken as a measure of the speech intelligibility.

The articulation score which is a relative quantity, is a function of a number of items which include the choice of sounds used (i.e. words, sentences, syllables etc.), the selection of testing personal and the manner in which the sounds are presented[3, 67].

The type of sounds used in articulation tests significantly affect the results. Words are more easily understood than syllables and sentences even more so. The intelligibility of sentences is determined by their meaning, context and rhythm[67] etc. These factors affect the scores and make the results difficult to analyse and interpret precisely.

Ideally a testing procedure should require little time, little or no special equipment and relatively untrained personnel[66]. Also it should be reliable and provide results which can be used to

assess the system function and the individual contribution from its major components.

6.2 DIAGNOSTIC RHYME TEST (DRT)

One test, called the Diagnostic Rhyme Test (DRT) has been developed to show small differences in a system and to offer a useful correlation between the parameters of the system and test scores. Also it has been demonstrated that it provides a reliable and economical measurement of speech intelligibility[68, 69]. The DRT is a method of evaluating a speech processing technique by determining the perceptual attributes of six English consonant phonemes. It measures the intelligibility and additional complex factors which are related to the peformance of the system under test.

The technique is to present a subject with a list of pairs of words. Each of which contain a single phonetic attribute but with different initial consonants. The listener is asked to identify which one of the words of the pair was spoken. The result of the listener's judgement is then used to determine the percentage score for the discriminabiltiy of each of these attributes. The percentage score is given by[68]:

$$D = \frac{R-W}{T} \times 100 \qquad 6.1$$

where D is the percentage of correct discriminations, R is the number of correct responses, W is the number of incorrect responses and T is the total number of responses.

The importance of this method in evaluating the analysis-synthesis technique, is in the relationship between different

perceptual attributes defined by different word lists and the frequency domain characteristics associated with these techniques. The perceptual attributes are related to the gain, voicing and spectral features of the speech[69]. This method has been used to evaluate different digital vocoding techniques[68, 69] and the evaluation results are used to improve the performance of these techniques.

6.3 HELIUM SPEECH EVALUATION TESTS

Reliable and consistent methods are needed to evaluate Helium speech unscramblers[35]. Intelligibility tests which give a single numerical rating can be used to compare different systems. These tests can also be used to improve the overall performance of the systems, by determining the responses of a descrambler to different speech attributes.

However, meaningful tests for the measurement of the intelligibility of a speech system working in specialised environments need a specific approach. They should be designed to measure the intelligibility with both listener and talker in the same environment. For example in noisy environments the intelligibility should be measured with the listener and talker subjected to the same degree of noise[67] since the vocal quality and listening ability becomes more significant in a noisy environment.

Another difficulty associated with Helium speech descrambler testing is the selection of the appropriate testing materials. The DRT, for example, needs a mimimum of four talkers and six

listeners[69] due to the sensitivity of DRT scores to individual listener and talker responses. However, these scores are much more sensitive to listener than to talker differences.

However, the words used for DRT should be read by at least two talkers in the same Helium oxygen environments. These recordings could be presented to a group of listeners who should ideally be in the same environment. Their average score would then be an indication of the system's performance.

6.4 SPECTRAL DISPLAY OF THE SPEECH SIGNAL

The important acoustic and perceptual features of the speech signal, such as formant, structure, voicing, stress, pitch, etc can be displayed using a device known as a sound spectrograph. This device provides a convenient way of displaying permanently the short time spectrum of a speech sample.

The spectrograph splits a short segment of a speech signal into equal frequency bands. It then measures the amplitude of the signal in each frequency band and displays them on a time-frequency plane. Special display papers are used in this instrument to enable the representation of different amplitude levels in the time-frequency plane. The darkness on this paper represents the intensity of different bands[3]. The bandwidth of the analysing filters determines the information displayed by the spectrograph. Filters with wide bandwidth provide better temporal resolution of the speech signal in the formants, whilst narrow bandwidths provides a frequency resolution adequate to resolve the harmonics of the voiced speech.

The sound spectrograph can be very useful in providing fast information on the characteristics of the system. The spectral features of which can be compared at the input and output of a system by comparing their spectrographs.

6.5 STEADY STATE SPECTRAL EVALUATION

The articulation tests, described in Section 6.2, require special testing facilities to evaluate Helium speech descramblers. Whilst spectral techniques require very specialised equipment. Unfortunately neither have been readily available to evaluate. Therefore the method adopted to evaluate its performance has been steady state spectral analysis.

A synthetic waveform is generated by simulating three speech formants, then a train of pulses is applied to their input to represent the voiced source whilst white noise at its input represents the unvoiced source.

Three second order bandpass filters of the switched capacitor type described in Chapter Five were used to simulate the speech formants. Their inputs were connected together whilst their output were summed to form a parallel synthesiser.

The frequency equaliser was loaded with data representing the Helium speech environment. This data was generated using a BASIC program (Appendix 2) and loaded into the frequency equaliser store using an EPROM emulator. The main processing time was chosen to be 32 milliseconds.

Also the amplitude equaliser was loaded with different envelope algorithms. These were generated using another BASIC program (Appendix 3). The input and output spectrum of the system were

displayed using a spectrum analyser. The analyser filters had a bandwidth of 300 Hz, and adjacent filters overlapped at their -4dB points.

6.5.1 VOICED CASE

The synthesiser was excited by impulses at a frequency of 120 Hz (average male pitch) and a duration of one hundred microseconds. The formant frequencies were 935 Hz, 2117 Hz and 2441 Hz respectively. The input spectrum for the generated signal is shown in Figure 6.1.

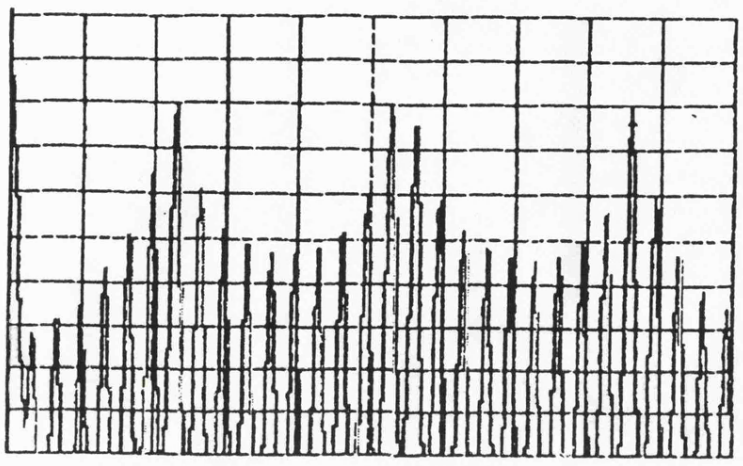
6.5.1 FORMANT COMPRESSION

The envelope spectrum for different frequency compression algorithms is shown in Figures 6.2a to 6.2e. In these cases no amplitude modification was included. The amplitude equaliser tables allow no modification of the signal envelope.

A comparison between these envelopes demonstrates the effect of different compression algorithms. A clearer conclusion could be derived by comparing the formant frequencies at the input of the sytem and at its output as shown in Table 6.1. The compression factors for different formants are shown in Table 6.2. It is clear that when the diving depth is large the formants shift nonlinearly (cases 4 and 5 in Tables 6.1, 6.2) whilst cases 2 and 3 represent linear compressions.

The fine resolution of the synthesised spectra are shown in Figure 6.3a to 6.3e. The spectrum in these figures has been drawn on an expanded scale. The frequency span for each case is $(\frac{1}{k_1})$ times the

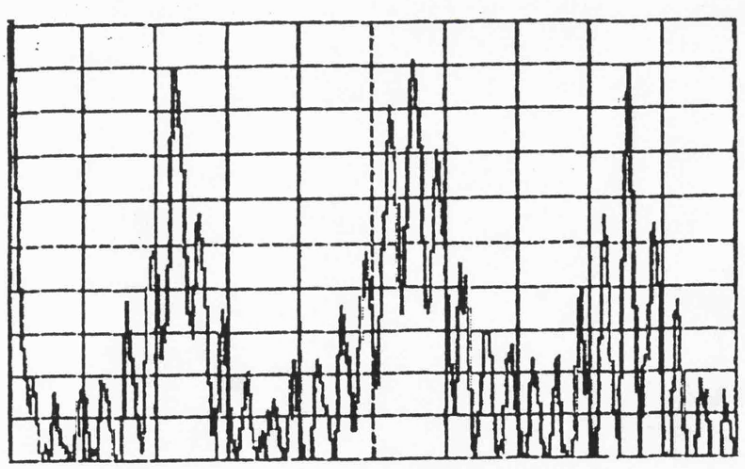
5dB/ div.



400 HZ/ div.

(a)

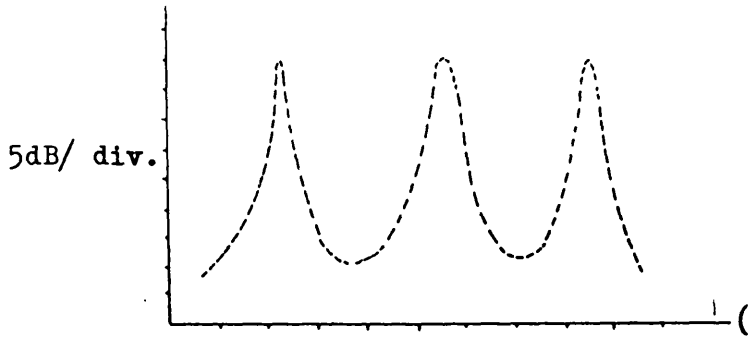
5dB/ div.



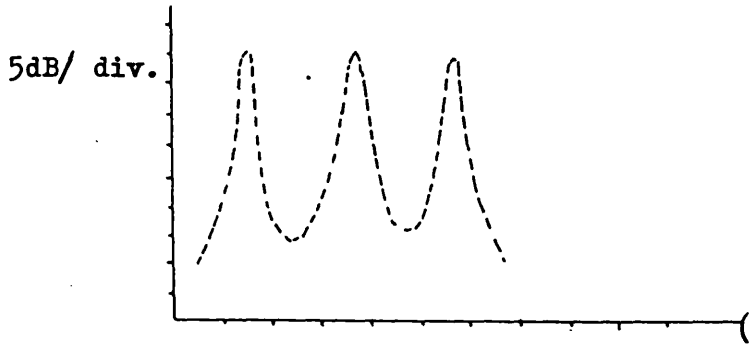
400 HZ/ div.

(b)

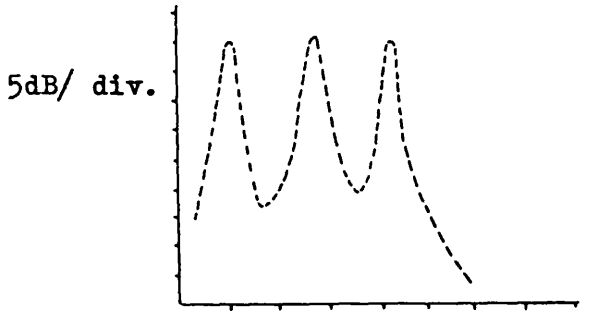
Figure 6.1 Spectrum of the Synthetic Signal at:
(a) the Input of the System;
(b) the Output of the System



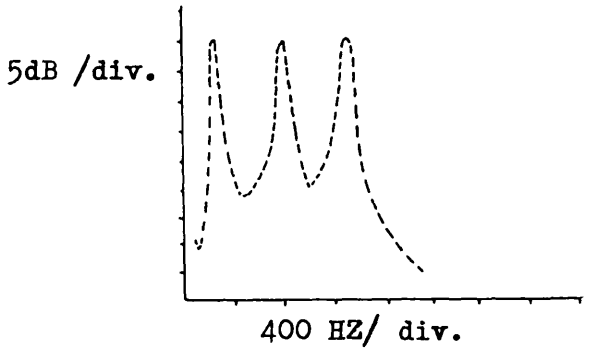
(a) Speed Ratio 1 ,
Depth 0 ft.



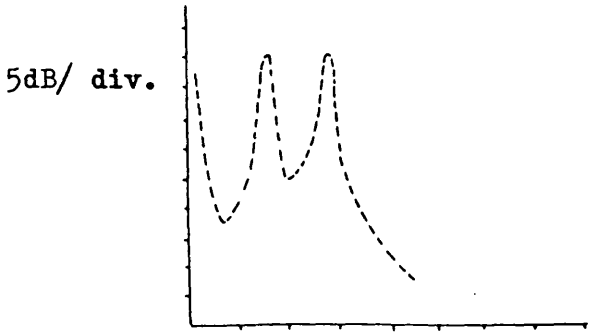
(b) Speed Ratio1.5,
Depth 50 ft.



(c) Speed Ratio 2 ,
Depth 200 ft.



(d) Speed Ratio 2.5,
Depth 400 ft.



(e) Speed Ratio 2.83,
Depth 1000 ft.

gure 6.2 Envelope Spectra of the Signal for Different Frequency Compression Algorithms.

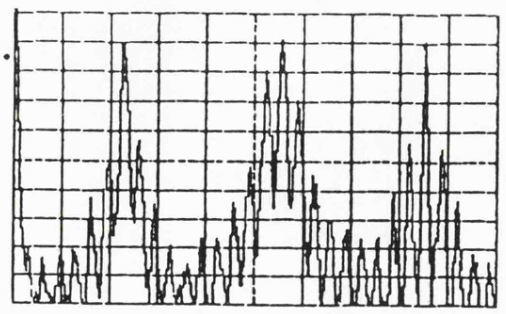
TABLE 6.1 Formant frequencies of the synthetic signal at the output of the system.

(a) formant's values (b) formant's ratio

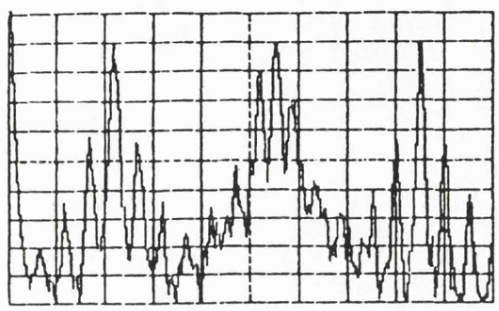
FORMANT	1	2	3	4	5
F_1 Hz	920	616	409	227	138
F_2 Hz	2110	1398	1031	798	733
F_3 Hz	3409	2276	1701	1345	1166

FORMANT RATIO	1	2	3	4	5
F_{1h}/F_{1a}	1.01	1.52	2.28	4.12	6.80
F_{2h}/F_{2a}	1.01	1.51	2.05	2.65	2.88
F_{3h}/F_{3a}	1.0	1.51	2.02	2.56	2.95

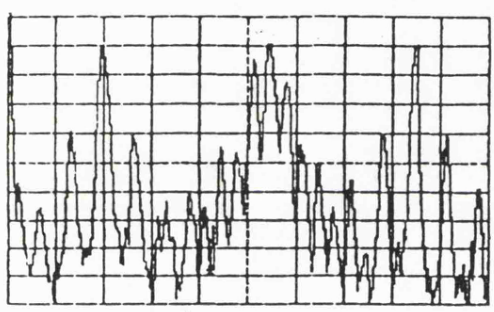
1) No frequency compression
2) speed ratio = 1.5, depth 50 feet
3) speed ratio = 2.0, depth 200 feet
4) speed ratio = 2.5, depth 400 feet
5) speed ratio = 2.83, depth = 1000 feet



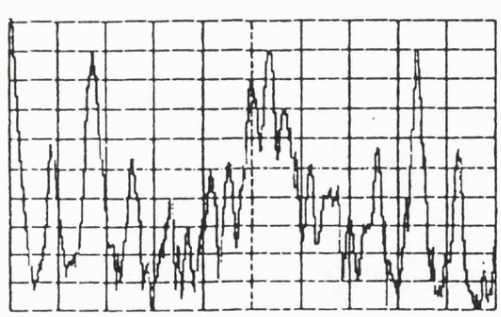
(a)Speed Ratio 1 , Depth 0 ft.



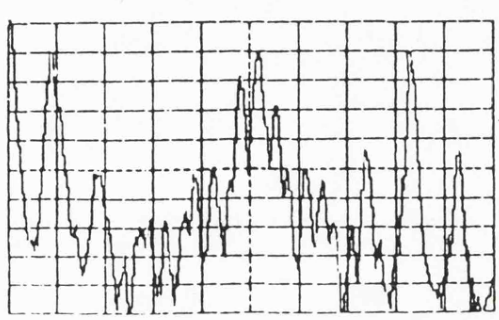
(b)Speed Ratio 1.5 ,Depth 50 ft.



(c)Speed Ratio 2 ,Depth 200 ft.



(d)Speed Ratio 2.5 ,Depth 400 ft.



(e)Speed Ratio 2.83 ,Depth 1000 ft.

Figure 6.3 Signal Spectra Drawn on Expanded Scale for Different Frequency Compression Algorithms.

scale of the input spectrum where K_1 is the linear expansion factor (speed ratio).

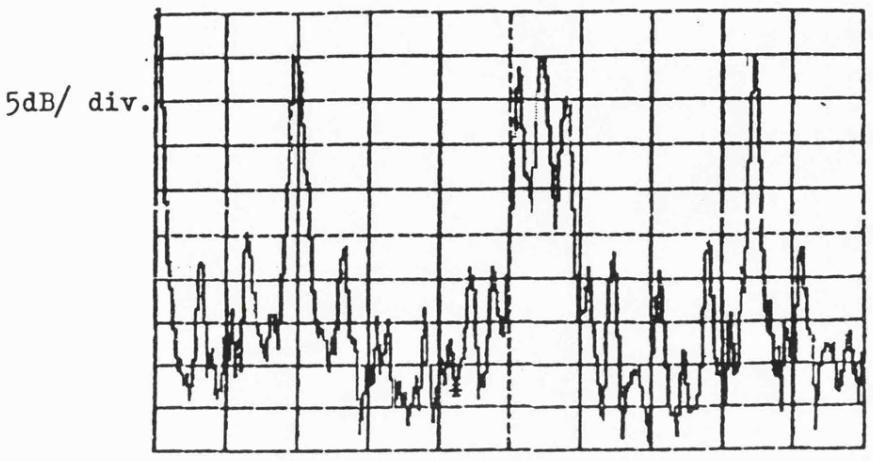
However, these spectra show that the fine resolution is preserved in the frequency compressed signal. This phenomena satisifies the specification of Helium speech descramblers, since the fine resolution is unchanged in both the normal and the Helium speech.

6.5.1.2 Amplitude Equalisation

As described in Chapters Two and Four Helium speech descramblers may require different types of algorithms to compensate for the frequency dependent attenuation of the Helium speech signal and formant bandwidths. The descramblers should be programmable to deal with amplitude equalisation.

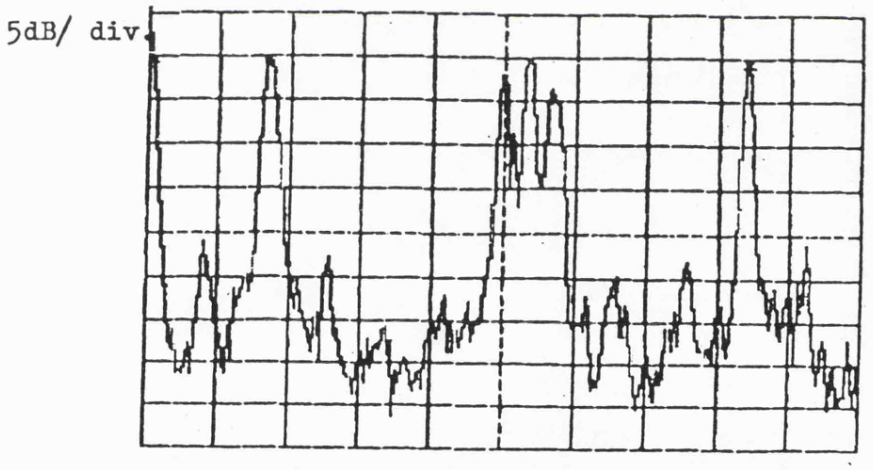
Two algorithms have been evaluated. For the rooting algorithm the envelope of the synthesised signal is rooted by a speed ratio as shown in Figure 6.4. Comparing the envelope spectrum of the signals with that shown in Figures 6.3c and 6.3d it is clear that the frequency components close to the formant frequencies are attenuated indicating a formant bandwidth compression. This has been demonstrated by plotting the envelope spectrum of the unmodified signal (Figure 6.5).

The second algorithm evaluated was designed to equalise frequency dependant attenuation. A synthetic signal was generated to have a spectrum shown in Figure 6.6a, and applied to the system. It was processed so that the amplitude of the third formant was amplified by 14dB with no actual frequency compression (Figure



200 HZ/ div.

(a)



160 HZ/ div.

(b)

Figure 6.4 Spectra of the Amplitude Equalised Signal for :

(a) Frequency Compression and Amplitude Rooting by Factor of 2.

(b) Frequency Compression and Amplitude Rooting by Factor of 2.5.

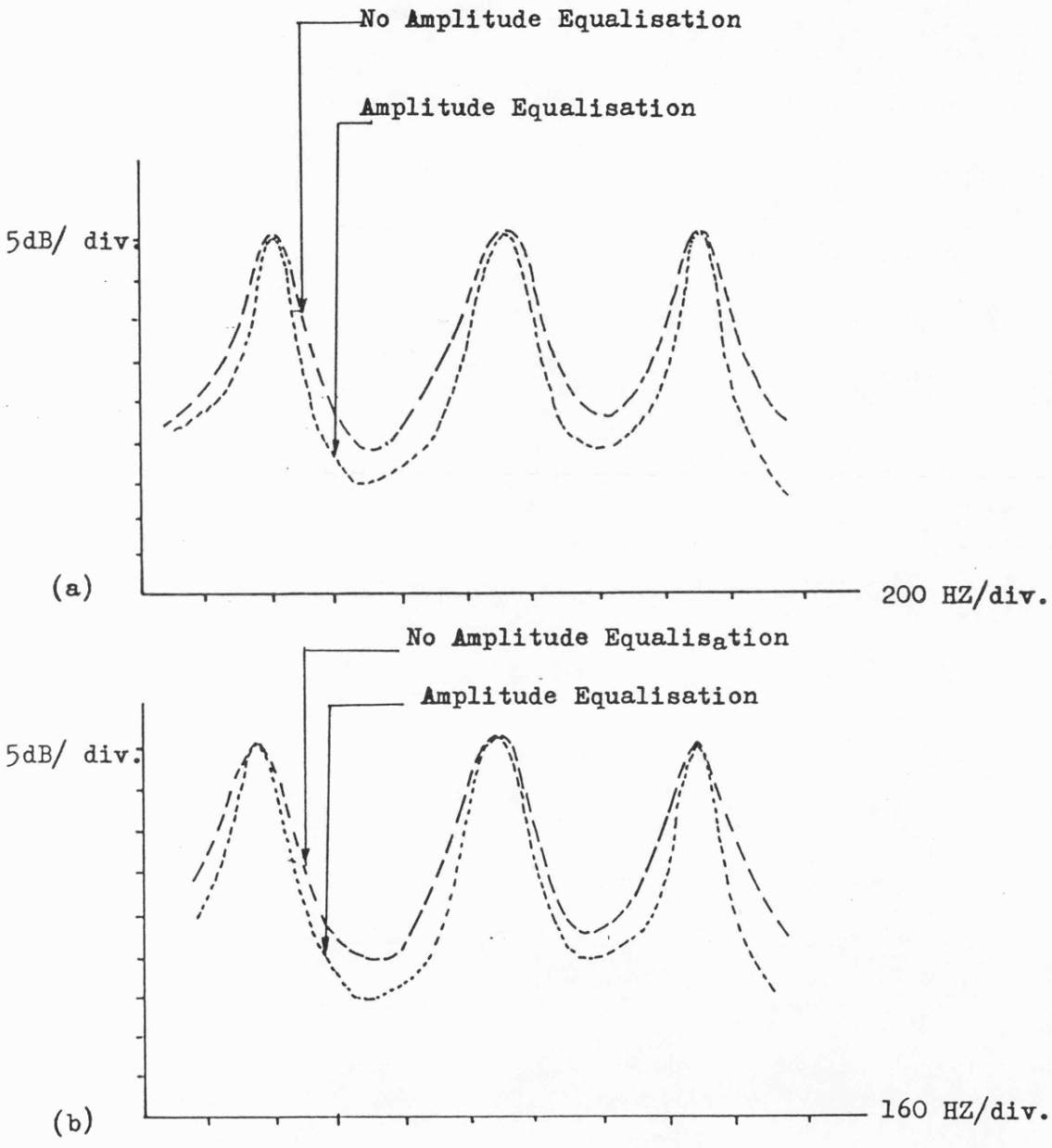
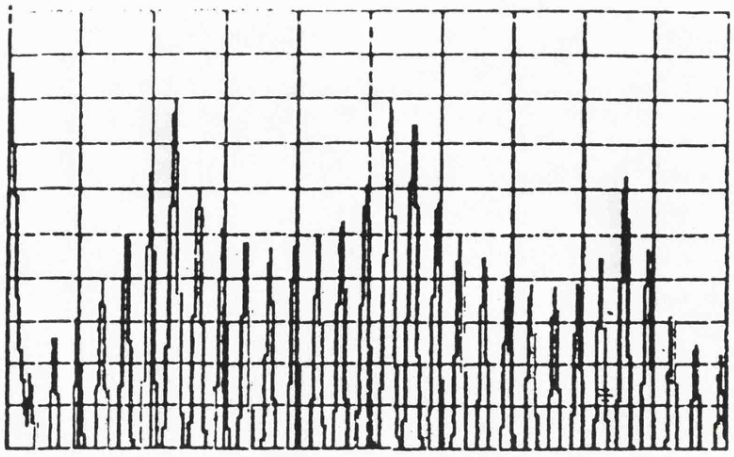
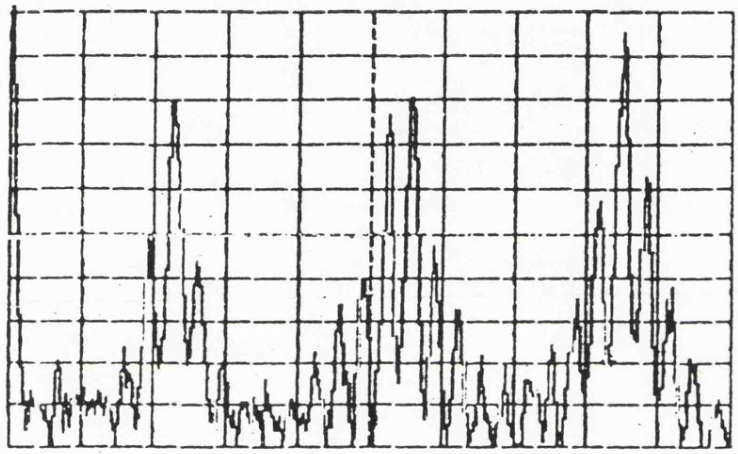


Figure 6.5 Envelope Spectra of the Amplitude Equalised Signal for : (a)Frequency Compression and Amplitude Rooting by factor of 2 ;

(b)Frequency Compression and Amplitude Rooting by factor of 2.5.





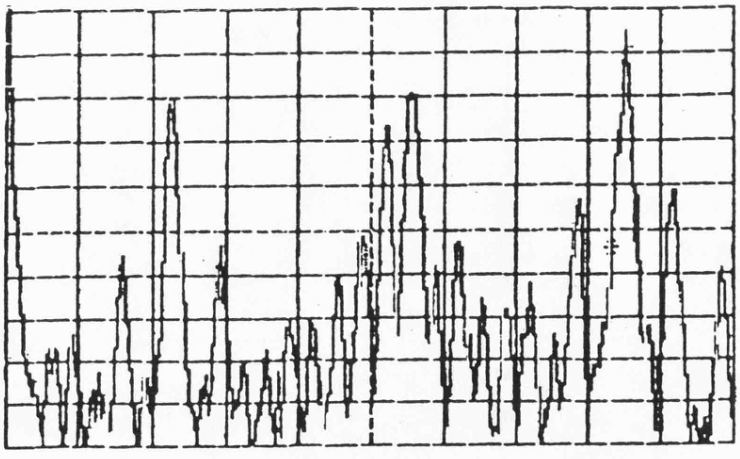


Figure 6.6 Spectra of the Amplitude Equalised Signal

(a) Input Spectra

(b) Output Spectra with no Frequency Compression- third Formant Amplification

(c) Output Spectra with Frequency Compression by factor of 2- third Formant Amplification

6.6b). Also frequency compression can be combined with amplitude compensation as shown in Figure 6.6c.

The processing of signals in individual channels by different amplitude laws can be realised by switching between different amplitude algorithms. However, combining amplitude rooting with straight amplification will reduce the overall bandwidth.

6.5.2 Unvoiced Case

A white noise generator was applied to the synthesiser to produce a synthetic signal with a spectrum shown in Figure 6.8.

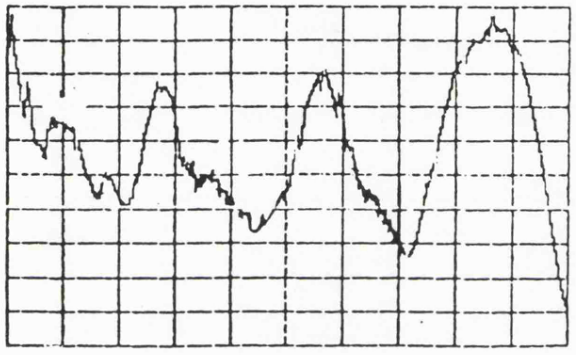
Three processing algorithms were applied to this signal as shown in Figures 6.7b and 6.7c. It is clear that the system compressed the spectrum of the unvoiced signal in the same manner as for the voiced signal. However, amplitude rooting of the noise signal reduces the amplitude of the low frequency components relative to the high frequency components. This effect is clearly potentially useful for Helium speech processing.

6.5.3 Narrow Band Case

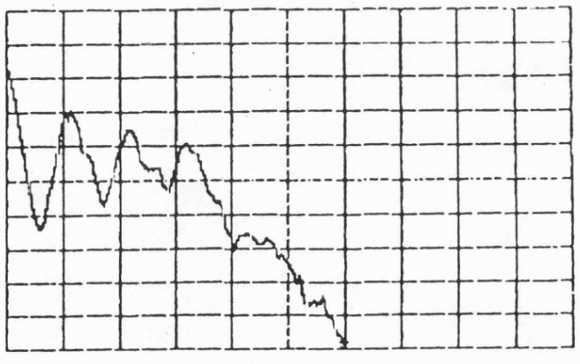
In certain speech processing techniques the analysis relies on processing the fine spectral structure of the speech. The parameters of the filters are chosen such that only one harmonic is present in their output. These signals are then processed in order to compress the frequencies of the speech signal.

To evaluate the system on a narrow band basis a harmonic generator with a fundamental frequency of 512 Hz was applied to its input. The spectrum of this signal is shown in Figure 6.8a whilst the output of the system is shown in Figure 6.9b. While figure 6.8c

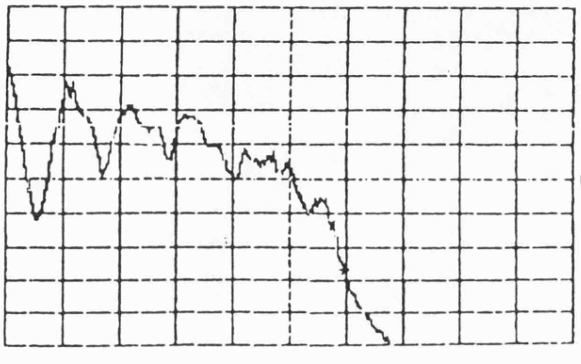
(a) Input Spectrum



(b) Output Spectrum with no Frequency Compression.

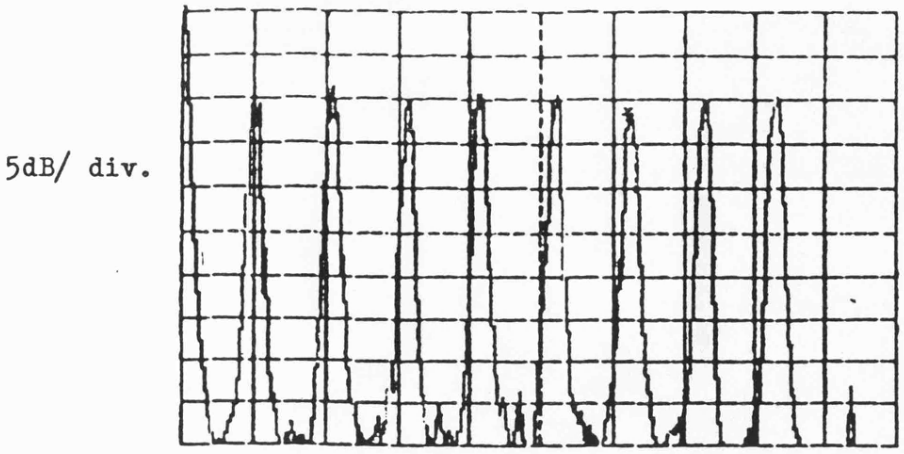


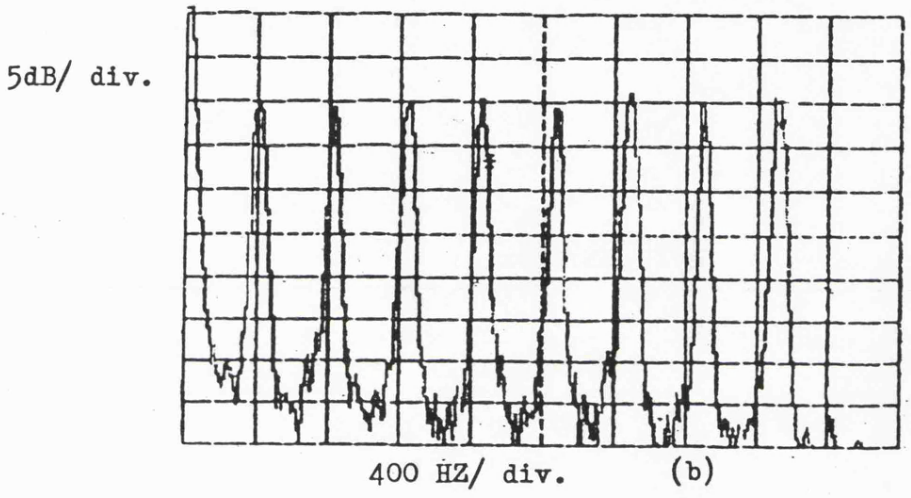
(c) Output Spectrum with Frequency Compression by factor of 2.5



(d) Output Spectrum with Frequency Compression and Amplitude Rooting by factor of 2.5

Figure 6.7 Noise Spectra at the Input and Output of the System.





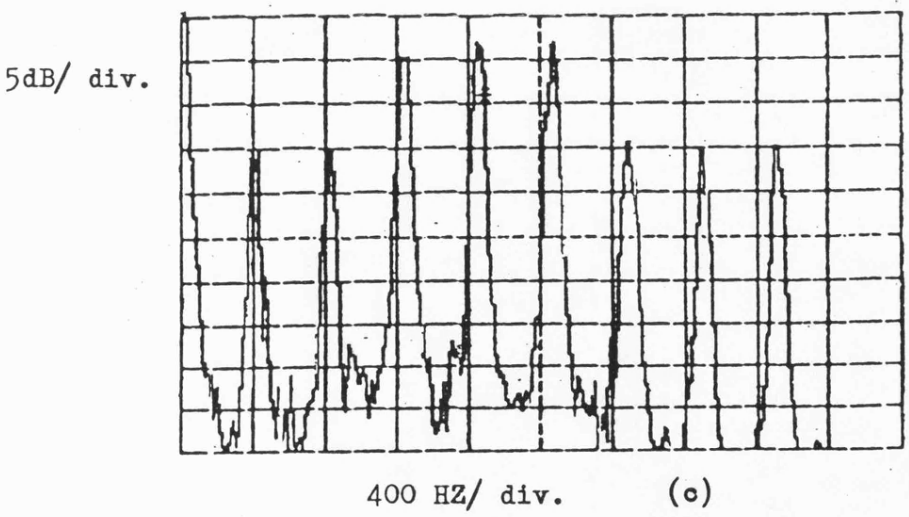


Figure 6.8 Spectra of the Harmonically Related Signal

(a) Input Spectra

(b) Output Spectra with no Frequency Compression

(c) Output Spectra With Frequency Compression by Factor of 2 and Amplitude Equalisation of the third , fourth , fifth Harmonics.

shows a compression of the frequencies by a factor of two, together with amplitude equalisation of the third, fourth and fifth harmonics. However, these tests show clearly the possible use of the system to compress harmonics of a signal when used as a narrow band analysis synthesis processing technique.

6.6 CONTINUOUS SPEECH TESTING

During the research programs a sample of Helium speech recorded at a depth of 150 metres became available. The percentage of Helium Oxygen mixture in which the speech was produced was unknown. Also there was a significant background noise level and the level of the speech fluctuated dramatically.

The frequency equaliser of the system was loaded with data derived from equation 2.16, after assuming that the Helium Oxygen mixture contained 90% Helium and 10% Oxygen (this estimation was derived from data given by Reference 35). The speed ratio (K_1) for this mixture would be 2.2 (from Table 2.1). The Helium speech from the tape was applied to the system through a 500Hz-6kHz bandpass filter, whilst the output from the system was recorded on another machine.

The recorded output was noisy and still difficult to understand. However, comparing the speech with the written text shows that an improvement in intelligibility had been achieved.

There could be many reasons for this result, other than those associated with the poor quality of the Helium speech recording and the unknown condition under which the recording was made. Firstly, the counting of the zero crossing as an estimation of the mean frequency spectrum leads to errors in the measurement of the formant

frequencies. In spite of the phase and amplitude symmetry in each analyser filter bandwidth, which is a necessary condition for the estimation of the mean frequency zero crossing rate,[63] practically, the measurement of zero crossing of the signal in a band will be greatly affected by the formant bandwidth. Damped sinusoid signals, which represent the formants in the time domain, and therefore decay rapidly reduce the number of detectable zero crossings. This effect leads to errors in the estimation of the formant frequency which will effect the speech intelligibility, especially when the frequency processing time is set high to maximise the frequency resolution. Another source of errors associated with zero crossing measurement is the dependence of the measurement on the residual noise level. The presence of noise increases the number of zero crossings in a band and randomizes the values of the formant frequencies with time. This can be reduced by adjusting the zero crossing threshold. However, increasing this threshold level in a noisy environment, as in Helium speech, will affect the formant frequency measurement.

A test was carried out to examine the effect of bandwidth on the speech intelligibility. The output frequencies from the synthesiser were set to specific values during the test. These frequencies were equal to the centre frequencies of the analysing filter divided by the speech ratio (K_1). The amplitudes of these signals were processed as before and the intelligibility of the processed speech was increased. Although the output speech had a synthetic quality because the formants have not been tracked dynamically, it demonstrated clearly that speech processed by this

method was highly affected by the way in which the formant frequencies were synthesised.

Also the error in the sampling rate of the signal envelopes was another reason for the poor performance of the unscrambler. The rapid decay of these envelopes requires a high sampling rate, to define clearly the start of each new pitch period and to recover the formant amplitudes. However, the sampling of these envelopes by this technique is limited by, the speed of the computer and the analogue to digital converter.

The converison time of the analogue converters was approximately ten microseconds. This conversion time requires the sample and hold circuit to store the sample of the envelope during the coversion interval. The minimum time required for sampling and converting a channel envelope into a digital word was therefore increased to some twenty microseconds and the minimum conversion time for all sixteen channels becaome 320 microseconds.

The speed limitation of the computer increases the time required for the envelope processing to one millisecond. If the average male pitch is considered to be 120Hz, then this processing rate could introudce an error of one millisecond into the pitch period value. This is equivalent to twenty five percent of the fundamental frequency of the male pitch. However, a one percent error in estimating the periods could produce unacceptable speech quality [7].

Further tests were carried out on normal speech to determine the effect of formant bandwidth on speech intelligibility. Normal speech was applied to the system and then resynthesised. The output from the synthesiser revealed certain of the system's

characteristics as described previously. This speech was both highly intelligible and had retained speaking identification. However, the naturalness was degraded since errors still existed in the pitch production and formant estimation.

The reason for the high intelligibility is that the formant bandwidths in normal speech is much smaller than in Helium speech which reduces the measurement errors in both formant frequencies and the speech envelopes.

Finally another source of error could be that the reduction in speech intelligibility at depths below 200 feet is not directly proportional to the Helium Oxygen mixture or the diving depth, but that below these depths there are other factors which affect intelligibility, such as the acoustic characteristics of the microphone in helmet.(25)

6.7 CONCLUSION

Speech processing systems require specialised intelligibility tests specially tailed for different processing techniques.

Intelligibility tests devised for analysis-synthesis techniques should ideally related the perceptual attributes of the speech to the frequency domain characteristics of the technique. These tests could then be used to determine speech intelligibility and also provide a means of improving the performance of the speech processing technique.

Tests which show acoustic features of the speech signal could provide useful information on the characteristics of the speech processing system.

Helium speech intelligibility tests need more specialised preparation than normal speech tests. These tests should then be carried out in similar diving environments.

The real time unscrambler based on the wide band analysis-synthesis technique has the ability to process most Helium speech distortions. However, this technique requires accurate measurements of the frequency envelope of the signals in different bands. The envelope of the signals has a major effect on those measurements.

CHAPTER SEVEN

CONCLUSIONS AND FURTHER WORK

7.1 CONCLUSION

Non linear compression of the frequencies of the speech signal is a useful method of correcting Helium speech distortions. Real time unscramblers based on wide band analysis-synthesis techniques for compressing the frequencies of the speech signal non linearly are a useful technique for processing Helium speech. It offers a generalised approach to the processing and analysis of Helium speech.

Useful information has emerged from the constructing and testing of the real time non linear frequency compression system. Although the tests were carried out over a limited range, they both indicated the ability of the system to non linearly compress the signal frequencies and also the flexibility of the system to correct frequency dependent attenuation. However, continuous speech tests indicated the dependence of the performance of the system on the formant envelopes.

The measurement of zero crossing rate as a method of obtaining the spectral mean of the signal in different bands gave unsatisfactory results for enhancing Helium speech due to rapid decay of the Helium speech envelope and the noise associated with the real deep diving.

The inaccuracy in the measurement of the envelope produced errros in determining the pitch frequency precisely which is another factor which degrades the speech quality due to the sensitivity of the ear to pitch.

Real time speech system with a software based correction algorithm provides flexibility for studying the effect of different speech parameters on its intelligibility.

It is possible to draw general conclusions from the study of Helium speech which might be useful for proposing or implementing Helium speech unscrambling systems. These conclusions will be used to suggest further work in the Helium speech field.

The Helium speech area is both speculative and confusing. There are uncertainties about almost all the Helium speech parameters. Even the well established relationship between formant frequencies in normal and Helium speech are now subject to doubt[26, 37]. The formant bandwidth and amplitude data is also doubtful. There are other factors which are not related to the physical characteristics of the speech production mechanism which affect the speech intelligibily beyond certain depths such as the microphone in the diver's helmet.

There are no reports of real time unscramblers which correct Helium speech in a totally satisfactory manner[26]. Unscramblers generally lack a system's approach and work only in a certain Helium speech environment.

There are no standard tests for evaluating the performance of Helium speech unscramblers or to compare alternative systems.

Finally Helium speech unscramblers should be designed only after a careful study of Helium speech characteristics. The information available at the present time is not adequate to produce a definitive specification. However, there is adequate information to provide an indication as to the area where investigations may most profitably be carried out.

7.2 FURTHER WORK

There are certain aspects of the wide band real time processor which has been constructed during this research program which would benefit from further investigation. For example, the effect of formant bandwidths on the mean frequency measurements and the possibility of measuring this frequency independent of formant bandwidth could greatly improve the unscrambler's performance. One possibility for which is the measurement of the average time between zero crossings in specific processing interval. However, it is possible to process the zero crossing rate in the computer prior to its equalisation in the system.

The sampling rate of the envelopecould be increased which would reduce the generated noise caused from inaccuracies in the determination of the speech pitch. This would require the use of a faster computer and analogue to digital converter.

For Helium speech the following specific areas would benefit from detailed investigation:

1) The change in formant bandwidths and their effect on Helium speech intelligibility.

2) The level of high frequency attenuation needs to be determined and the frequency over which this attenuation occurs should also be defined. The relation between this attenuation and the diving environments should be available to the designer of Helium speech unscramblers.

3) Work is required to determine the relationship between the formant frequencies of Helium speech and those of normal speech,

particularly after the recent speculation[26, 37] concerning this apparently known relation.

4) Thorough study is required to determine precisely the change in the source spectrum in the Helium Oxygen environment.

5) The reasons for the reduced rate of deterioration of intelligibility with depth beyond a critical depth should be investigated since it significantly affects the design specification of an unscrambler.

6) Special intelligibility tests are clearly required to evaluate Helium speech unscramblers. These might be derivative of conventional speech intelligibility tests. Test elements should be carefully chosen such that they reflect the important Helium speech parameters in the frequency domain. Also the precise environments in which these tests are carried should be carefully defined.

7) Simulating Helium speech clearly provides an invaluable facility for the design of any conceivable unscrambler system.

REFERENCES

1. Rabiner, L.R., Schafer, R.W., "Digital processing of speech signals". Prentice Hall, New York, 1980.

2. Stremler, G.F., "Introduction to communication sytem". Second Edition, AddisonWesley, 1982.

3. Flanagan, J.L., "Speech analysis synthesis and perception", Spring Verlag, 1972.

4. Holmes, J.N., "A survey of methods for digitally encoding speech signal". Journ. of I.E.R.E., Vol. 52, pp. 267-276. 1982.

5. Dudley, H., "The Carrier nature of speech". Bell System Tech. J., 19, pp 495-511, 1940.

6. Lawrence, W., "The synthesis of speech from signals which have a low information rate". In Speech Synthesis, Dowden, Hutchinson and Ross, pp 234-243, 1973.

7. Schroeder, M.R., "Vocoders: Analysis and synthesis of speech". Proc. IEEE, Vol. 54, pp 720-734, May 1966.

8. David, E.E., Schroeder, M.R., Logan, B.E. and Prestigiacomo, A.J., "Voice excited vocdoer for practical speech bandwidth reduction". I.R.E. Tran. Information Theory, Vol. IT8, pp S 101-105, Sep. 1962.

9. Marcou, P. and Daguet, J., "New methods of speech transmission". Proc. 3rd Symp. on Information Theory. London, England, 1955, Information Theory, C. Cherry, Ed. London, England: Butterworths, pp 231-244, 1956.

10. Bogert, B.P. "The Vocab.A-two-to-one speech bandwidth reduction system". J. Acoust. Soc. Am., vol. 28, No. 3, pp 399-404, May 1956.

11. Daguet, J.L. "Speech compression codimex system", IEEE Trans. on Audio, Vol. AU11, pp 63-71, March April, 1963.

12. Schroeder, M.R., Flanagan, J.L., Lundry, E.A. Bandwidth compression of speech by analytic signal rooting". Proc. IEEE, Vol. 55, No. 3. PP 396-401, march 1967.

13. Bogner, R.E. "Frequency division in speech bandwidth reduction". IEEE Trans. on Communication Technology, Vol. Com. 13, No. 4, pp 438-451, Dec, 1965.

14. Bogner, R.E. and Flanagan, J.L. "Frequency Multiplication of speech signals". IEEE Trans. on Audio and Electronics, pp 202-208, Sept, 1969.

15. Flanagan, J.L. and Golden, R.M. "Phase Vocoder" Bell Syst. Tech. Vol 45, pp 1493±1509,Nov, 1966.

16. Bartholamew, C.A., "Commercial and military diving in the United States today". Navy Experimental Diving Unit Reports, Panama city, Florida, 1979.

17. Drubaker, D.J., Wurst, J.W. "Spectrographic analysis of diver's speech during decompression". Jour. Acous. soc. Am., Vol. 43, No. 4 pp 798-802, Apr. 1968.

18. Richards, M., "Helium speech enhancement using short time fourier transform", IEEE trans. on acoustic speech and signal processing, Vol. Assp 30, No. 6, pp 841-853, Dec 1982.

19. Vestrheim, M., Hatlestad, S., Blecher, E. and Sleithei, K. "Deep ex81 diver communication" Norwegian Underwater Technology Centre, Report 1782, Jan 1982.

20. Beil, R.G. "Frequency analysis of vowels produced in a Helium-rich atmosphere". Jour. Acous. Soc. Am., Vol. 34, No. 3, pp 347-349,March 1962.

21. Maclean, D.J. "Analysis of speech in HeliumOxygen mixture under pressure". Jour. Acous. Soc. Am., Vol. 40, No. 3, pp 347-349, March 1962.

22. Fant, B., Sonesson, B. "speech at high ambient air pressure". Roy. Inst. Tech., Stockholm, Sweden, Quat., Status Rep., Vol. 2/1964, pp 921. Jan. 1964.

23. Tanaka, R., Nakatsui, M., Suzuki, J. "Formant frequency under high ambient pressures'. Jour. Radio Research Lab., Vol. 21, No. 105, pp. 261-267,1974.

24. Tanaka, R., Nakatsui, M., Takasugi, T. and Suzuki, J. "Source characteristics of speech produced under high ambient pressures". Jour. Radio Research Lab., Vol. 21, No. Lo5, pp 269-273, 1974.

25. Rothman, H.B., Hollier, R.G., Lambetsen, C.J. "Speech intelligibility at high HeliumOxygen pressure". Under sea Biomedical Research, Vol. 7, No. 4, pp 265-274,Dec. 1980.

26. Belcher, E.O. "Formant frequencies, bandwiths and Q's in Helium Speech". Jour. Acous, Soc. Am., Vol. 74(2), pp 428-432, Aug. 1983.

27. Holywell, K., Harvey, G. "Helium Speech" Jour. Acous. Soc. Am., Vol. 36, pp 210-211,Jan. 1964.

28. Copel, M. "Helium voice unscrambling". IEEE on Audio Electroacoustics, Vol. 4, No. 3, pp 122-126,Sep, 1966.

29. Stover, W.R. "Technique for correcting Helium speech distortion". Jour. Acous. Soc. Am., Vol. 41, No. 1, pp 7074, 1967.

30. Roworth, D.A.A., "A practical processor for Helium speech". Naval Underwater Engineering Symposium, 13th May 1969.

31. Giordano, T.A., Rothman, H.B., Hollien, H. "Helium speech unscrambler: A critical review of the state of art". IEEE Tran. on Audio and Electroacoustic, Vol. Au. 21, No. 5, pp 436-444, Oct 1973.

32. Suzuki, J., Nakatsui, M. "Translation of Helium speech by splicing of autocorrelation function". Jour. Radio Research Lab., Vol. 23, No. 111, pp 229234, July 1976.

33. Suzuki, J., Nakatsui, M., Takasugi T., and Tanaka, R. "Translation of the speech by the method of segmentation, Partialrejection and expansion". Jour. Radio Research Lab., Vol 24, No. 113, pp 116, March 1977.

34. Jack, M.A., Miline, A.d., Virr, L.E. "Compact Helium speech unscrambler using charge transfer devices". Electronic letters, Vol. 5, No. 18, pp. 548-550 30th Aug. 1979.

35. Richards, M.A. "Helium speech enhancement using the short time fourier transform". Ph.D Thesis, School of Elect. Eng. Georgian Institute of Tech. 1982.

36. Nakatsui, M. "Comments on Helium speech insight into speech event needed". IEEE Tran. on Acous., Speech and Signal Processing, pp. 472-473 Dec. 1974.

37. Beet, S.W., Goodyear, C.C., "Speech formant shifts in Hyperbaric Helium". Spring Conference Acoust. 84, 9-12, April, 1984, University College of Swansea.

38. Jack, M.A., Duncan, G. "The Helium speech effect and electronic techniques for enhancing intelligibility in a HeliumOxygen environment". I.R.E.E., Vol. 52, No. 5, pp 211-223, May 1982.

39. Beet, S.W., Goodyear, C.C., "Helium speech processing using linear prediction". Electronic letters, Vol. 29, No. 11, pp. 408-409 May 1983.

40. Fant, G. "Acoustic Theory of Speech Production". Mouton and Co. SGravehange 1960.

41. Flanagan, J.L. "Note on the deisgn of "Terminal Analogue" speech synthesiser". Jour. Acous. soc. Am., Vol. 29, pp 306-310, 1957.

42. Fant, G., Liljencrants, L. "Acoustic analysis and synthesis of speech with application to Swedish" Ericson Technics, 15, No., 1, pp 3-108, 1959.

43. Witten, I.H. "Principle of computer speech", Academic Press, 1982.

44. Lim, J.S. and Oppenheim, A.V. "Enhancement and bandwdith compression of noisy speech". Proc. IEEE, Vol. 67, No. 12, pp 1586-1604. Dec. 1979.

45. Flanagan, J.L., Christensen, S.W. "Computer studies on parametric coding of speech spectra". Jour. Acous. Soc. Am., Vol. 68 (2), pp 420-430, Aug 1980.

46. Schroeder, M.R. "Models of Hearing" Proc. IEEE, Vol. 63, No. 9, pp 1332-1350, Sept 1975.

47. Scharf, F. "Critical Bands". In Tobias, J.V. (Ed.) "Foundations of modern auditory theory". (Academic Press, 1970, pp. 157-202.

48. Plomp, R., "The ear as a frequency Analyser". Jour. of Acous. Soc. Am., Vol. 36, No. 9, pp 1628-1636, Sept. 1964.

49. Oppenheim, A.V. "Applications of digital signal processing" Prentice Hall, 1978.

50. Pollack, I. "The information of Elementary auditory displays". Jour. of Acous. Soc. Am., Vol. 24, No. 6, pp 745-749, Nov. 1952.

51. Rossing, T.d. "The science of sound". Addison-Wesley, 1982.

52. Flanagan, J.L. "Estimate of the maximum precision necessary in quantizing certain dimensions of vowel sounds". Jour. Acoust. Soc. Am., Vol. 29, pp 533-534, Apr. 1957.

53. Quick, R.F. "Helium speech translation using Homomorphic Techiques". USAF Cambridge Res. Labs. Cambridge, Mass., Phys., Sci. Res. Paper 425, July 1970.

54. Flower, R.A., Gerstman, L.J. "Correction of Helium speech distortions by real time electronic processing". IEEE Tran. on Summ. Tech., pp 362-364, June 1971.

55. Suzuki, H., Ooyama, G., Kido, K. "Analysis conversion-synthesis system for improving naturalnes and intelligibility of speech at high pressure Helium gas mixture". Speech Comm. Seminar, Stockholm, pp 97-105, Aug. 1-3 1974.

56. Morris, R.J., Cook, W.J. "Helium speech converter operating in the time domain" Oceanology International, pp 482-484, 1972.

57. Golden, R.M. "Improving naturalness and intelligibility of Helium-Oxygen speech uisng Vocoder Techniques". Jour. Acous. Soc. Am., Vol. 40, No. 3 pp 621-627, 1966.

58. Zucher, J.F. "Voice transcoder". British Patent Specification 1561918, 5th March, 1980.

59. Malah, D., Flanagan, J.L. "Frequency scaling of speech signals by transform techniques" Bell System Technical Jour., Vol. 60, No. 9, pp 2107-2156, May 1981.

60. Takasungi, T., Suzuki, J. "Translation of Helium speech by use of analytical signal". Jour. radio Res. Lab's Vol. 21, pp 61-69, 1974.

61. Makhoul, J. "Methods for nonlinear spectral distortion of speech signal" Proc. IEEE Int. Conf. on Acous., Speech, and Signal Processing. pp 87-90, April 1976.

62. Baghdady, E.J. "Lectures on Communication System Theory" McGraw Hill Book Company, 1961.

63. Morris, L.R. "The role of zero crossings in speech recognition and processing". PhD Thesis, Dept. of Elect. Eng., Imperial coilege of Science and Tech. University of London, 1970.

64. Gant, G., Liljencrants, J. "How to define formant level: A matimatical model of voiced sounds" Speech Transmission Lab. Royal Inst. of Tech., STL/QPSR-2, pp 1-9, 1962.

65. Williams, A. "Electronics of filter design handbook". McGraw Hill, 1981, p. 257.

66. House, A.S., Williams, Hecker, M.H.L. and Kryter, K.D. "Articulation-testing methods: consonontal differentiation wih a closed-response set". The Jour. of Acous Soc. Am., Vol. 37, No. 1. pp 158-166, Jan 1965.

67. Beranek, L.L. "Acoustic measurements". John Wiley, 1967.

68. Voiers, W.D. "The present state of digital vocoding technique: A diagnostic evaluation". IEEE Tran. on Audio and Electro Acous., Vol Au-16, No. 2, pp 275-270, June 1968.

69. Wong, D.Y., and Markel, J.D. "An intelligibility evaluation of several linear prediction vocoder modifications". IEEE Tran. on Acous., Speech, and Signal Processing, Vol. Ass p-26, No. 5, pp 424-435, Oct, 1978.

APPENDIX I Data for weighting the envelope of the signal

level dB	Location Address	Output to Amplitude Weighting Circuit
0	255	0
1	227	1
2	202	2
3	180	3
4	160	4
5	143	5
6	127	8
7	113	9
8	101	10
9	90	11
10	80	12
11	71	13
12	64	16
13	57	17
14	50	18
15	45	19
16	40	20
17	36	21
18	32	24
19	28	25
20	25	26
21	22	27
22	20	28
23	18	29
24	16	32
25	14	33
26	12	34
27	11	35
28	10	36
29	9	37
30	8	40
31	7	41
32	6	42
33	5	43
34	5	44
35	4	45
36	4	48
37	3	49
38	3	50
39	2	51
40	2	52
41	2	53
42	2	56
43	1	57
44	1	58
45	1	59
46	1	60
47	1	61

APPENDIX 2

```
5 REM....TO.GENERATE..FREQUENCY..EQUALISER.DATA
10  X=&H8000
50  REM.TO.LOAD.EPROM1.USE.SUB.470
60  REM..TO..LOAD..EPROM..2...USE..SUB..610
90  POKEX,I:X=X+1:NEXT
110  REM...GOSUB700
130  T=.032
135  REM..K1-SOUND VELOCITY RATIO..G-GAMMA RATIO.D..DEPTH
140  REM.T-PROCESSING.TIME
150  FORL=0 TO 7:READ K1,G,D
170  Z1M=T*7800
180  REM.Z1M-MAXIMUM.ZERO.CROSSINGS
190  IF Z1M>255 THEN 210 ELSE 230
210  Z1M=255
230  FOR Z=0TO Z1M:F=Z/T:P=1+.0294*D:F0=SQR(G*P-K1*K1)*F1)
250  IF F < F0 THEN 270 ELSE 290
270  I=0 :GOTO 330
290  I=SQR((F*F-F0*F0)/(K1*K1))
300  I=INT((I/6.1)*1000+.5)/1000):GOSUB470
330  POKEX,I:X=X+1:NEXT
350  IF Z <255  THEN 370  ELSE 410
370  FOR Y=Z TO255 :I=0:POKEX,I:X=X+1
390  NEXT
430  NEXT
450  STOP
```

```
470  IF I > 255 THEN 490 ELSE 590
490  I=I-256:IF I > 255 THEN 510 ELSE 590
510  I=I-256:IF I > 255 THEN 530 ELSE 590
530  I=I-256:IF I > 255 THEN 550 ELSE 590
550  I=I-256:IF I > 255 THEN 570 ELSE 590
570  I=I-256
590  RETURN
610  IF I > 255 THEN 630 ELSE 650
630  I=I-256:IF I > 255 THEN 670 ELSE 710
650  I=0:RETURN
670  I=I-256:IF I>255 THEN 730 ELSE 690
690  I=2:RETURN
710  I=1:RETURN
730  I=I-256:IF I > 255 THEN  770 ELSE 750
750  I=3:RETURN
770  I=I-256:IF I > 255 THEN 810 ELSE 790
790  I=4:RETURN
810  RETURN
830  DATA1,1,0,1.5,1.1,50,2,1.14,200,2.5,1.18,400,2.83
       ,1.19,1000,2,1.14,0,2.5,1.18,0,0,0,0
```

APPENDIX 3

```
10.AMPLITUDE.EQUALISATION.DATA
20 REM..LOAD..OF.APPENDIX.1.INTO.COMP.RAM
30 X=&H8000
40 REM..A-AMPLITUDE ROOTING FACTOR
50 REM..B-THRESHOLD.LEVEL
60 J=0 TO 5: READ A,B
70 FOR I=0 TO 255
80 IF Y > B THEN 90 ELSE 100
90 Y=6 :GOTO 160
100 GOSUB 350
110 Y=INT((Y*1000+.5)/1000))
120 IF Y > 255 THEN Y=255
130 IF  Y < 0 THEN Y=0
140 Y=PEEK(&H7FFF-Y)
150 IF Y > B THEN Y=6
160 POKE X,Y:X=X+1:NEXT:NEXT
170 REM...A-AMPLIFICATION FACTOR
180 FOR J=0 TO 1:READ A,B
190 FOR I=0 TO255
200 Y=PEEK(&H7FFF-I)
210 IF  Y > B THEN 220 ELSE 230
220 Y=6: GOTO 260
230 GOSUB 400
240 Y=INT(Y*1000+.5)/1000))
250 IF Y > 255 THEN Y=255
260 IF Y < 255 THEN Y=0
```

```
270 Y=PEEK(&H7FFF-Y)
280 IF Y >B THEN Y=6
290 POKE X,Y: X=X+1
300 NEXT:NEXT
350 C=(255[(1/A)):Y=(I[(1/A)*(255/C))
360 RETURN
400 C=10[(A/20):Y=I*C
410 RETURN
420 DATA 1,48,1.5,48,2,48,2.5,48,2.83,48
430 DATA 10,48,14,48
```